OPEN ACCESS AND THE PUBLIC DOMAIN IN DIGITAL DATA AND INFORMATION FOR SCIENCE

PROCEEDINGS OF AN INTERNATIONAL SYMPOSIUM

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

Open Access and the Public Domain in Digital Data and Information for Science

PROCEEDINGS OF AN INTERNATIONAL SYMPOSIUM

Julie M. Esanu and Paul F. Uhlir, Editors

U.S. National Committee for CODATA

Board on International Scientific Organizations

Policy and Global Affairs Division

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

Support for this project was provided by the John D. and Catherine T. MacArthur Foundation (under Grant No. 02-74944-000-GEN) and the National Oceanic and Atmospheric Administration/National Weather Service (under an unnumbered purchase order). Additional support was provided by the Committee on Data for Science and Technology, the International Council for Science, and the United Nations Educational, Scientific and Cultural Organization. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number 0-309-09145-4 (Book)
International Standard Book Number 0-309-53007-5 (PDF)

Copies of this report are available from the Board on International Scientific Organizations, 500 Fifth Street, N.W., Washington, DC 20001; 202-334-2807; Internet, http://www7.nationalacademies.org/biso/.

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

Printed in the United States of America

ADDENDUM

Our deepest apologies. Shuichi Iwata's biography was inadvertently omitted from Appendix B, the list of Biographical Summaries of Symposium Speakers and Steering Committee Members.

Shuichi Iwata's biography is listed below:

Shuichi Iwata is professor of data science and environmental engineering at the University of Tokyo and president of CODATA, the Committee on Data for Science and Technology, a term he holds until 2006. Dr. Iwata received his doctorate in nuclear engineering from the University of Tokyo in 1975. He has served in various capacities at the University of Tokyo, including lecturer, associate professor, and head of the Metallurgical Division (1978-1981) of the Engineering Research Institute; associate professor of nuclear fuels and materials (1981-1991); professor of materials design (1991-1992), Department of Nuclear Engineering; professor of design science (1992-2002), director (1997-2000), and professor of life cycle engineering (2002-2003) at the RACE (Research into Artifacts, Center for Engineering) Center; and professor of design science of materials (2003-2004), Quantum Engineering and Systems Science, School of Engineering. He served as a guest researcher in FIZ-Karlsruhe, Germany from October 1985 to October 1986. His work includes researoh on design science of materials and engineering products, nuclear fuels, and materials and materials databases. He has also served as project leader and coordinator in the fields of materials databases and materials design. Dr. Iwata serves a chairman of JSPS 122 Committee and member of SCJ Liaison Committee. He is a member of the academic societies for the Japan Institute of Metals, the Iron and Steel Institute of Japan, Japan Society of Energy and Resources, the Physical Society of Japan, Information Processing Society of Japan, and the Atomic Energy Society of Japan. He received the Promotion of Science and Technology Information Award from JST in 1998, a paper award from the Japan Institute of Metals in 1999, and the GIW Best Paper Award in 2003.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

Preface

Data and information produced by government-funded, public-interest science is a global public good caught between two different trends. On the one hand, the Internet provides valuable new opportunities for overcoming geographic limitations and the promise of unprecedented open access to public information for research on a global basis. The synergistic aspects of the availability and access to such information result in a broad range of positive externalities and network effects that increase exponentially with the addition of new Internet users. On the other hand, there are growing restrictions on the availability and use of public data and information arising from the privatization and commercialization of such sources. This countervailing trend undermines the traditional scientific cooperative and sharing ethos. It diminishes the public domain and open access to such global public goods and leads to a host of lost opportunity costs at both the national and international levels.

While there has been a great deal of focus on new commercial opportunities with digital information and on increased intellectual property rights, comparatively little attention has been devoted to the importance of maintaining open access to the source of upstream scientific—and other—data and information produced in the public domain for the benefit of all downstream users, or to the imperative to balance the public and private interests. The question is how to preserve and promote access to and sharing of such public scientific resources without unduly restricting new opportunities for commerce or the rights of authors. Conversely, how should commercial activities in the private sector be promoted without significantly compromising the availability of data and information in the public domain or through open access for global public good purposes?

The recent pressures on both public-domain and open-access information—scientific and otherwise—have resulted from a variety of legal, economic, and technological factors. New and revised laws have broadened, deepened, and lengthened the scope of intellectual property and neighboring rights in data and information, substantially redefining and limiting the public domain. National security concerns also are constraining the scope of government data and information that can be made publicly available. Economic pressures on both government and university producers of data and information similarly have narrowed the scope of such information placed in the public domain, with resulting access and use restrictions on resources that were previously openly available to researchers, educators, and others. Advances in digital rights management technologies for enforcing proprietary rights in various information products are posing some of the greatest potential restrictions on the public domain and open access to data and information.

Nevertheless, some well-established mechanisms for preserving public-domain or open-access data and information—such as public archives and data centers, together with ever-increasing numbers of open Web sites—exist

in the government, academic, and not-for-profit sectors. Very innovative institutional and legal models for making available digital scientific data and information resources in the public domain or through open-access provisions are now being developed by different groups in the scientific, library, and legal communities in many countries.

To address these issues the International Council for Science (ICSU), the United Nations Educational, Scientific and Cultural Organization (UNESCO), the U.S. National Academies, the Committee on Data for Science and Technology (CODATA), and the International Council for Scientific and Technical Information (ICSTI) jointly organized a major international symposium on "Open Access and the Public Domain in Digital Data and Information for Science." This symposium, which was held on March 10-11, 2003, at UNESCO headquarters in Paris, brought together policy experts and managers from the government and academic sectors in both developed and developing countries to (1) describe the role, value, and limits that the public domain and open access to digital data and information have in the context of international research; (2) identify and analyze the various legal, economic, and technological pressures on the public domain in digital data and information, and their potential effects on international research; and (3) review the existing and proposed approaches for preserving and promoting the public domain and open access to scientific and technical data and information on a global basis, with particular attention to the needs of developing countries.

The symposium, along with the Workshop on Science in the Information Society, which was organized by ICSU, CODATA, and UNESCO and held on March 12, 2003, at UNESCO headquarters in Paris,[1] also helped to identify and analyze important issues for follow up by the ICSU family of organizations. The results of this subsequent workshop were summarized by ICSU and used to provide scientific community input for the development of a Declaration of Principles and Plan of Action in preparation for the World Summit on the Information Society.

The symposium was organized into six sessions, each introduced by a moderator and then followed by several invited presentations. The first session focused on the legal, economic, and technological framework for open access and public domain in digital data and information for science. The following sessions explored the opportunities and challenges of open-access and public-domain scientific information in developing countries in the areas of data and information in the public health and environmental sectors, the basic sciences, and higher education. The summary concluded with a discussion of innovative models for public-domain production of open access to scientific and technical data and information, with a focus on examples of new initiatives for promoting open access in developing countries.

Different aspects of the issues discussed in this symposium have already been addressed in some detail in several reports published by the National Academies.[2] More specifically, the Office of International Scientific and Technical Information Programs (ISTIP) recently convened a "Symposium on the Role of Scientific and Technical Data and Information in the Public Domain."[3] The March symposium built on the results of the ISTIP symposium.

The results from these studies and activities provided a solid foundation for holding in-depth discussions of the issues relating to public domain and open access in digital data and information produced or used by public interest science. Although these previous works addressed various aspects of these issues in detail, none provided an international focus and forum at which representatives of public and private interest groups and experts could discuss them in a public venue. Over 150 experts attended the meeting (see the list of participants in Appendix C).

This publication presents the proceedings at the symposium. The speakers' remarks were taped and transcribed, and in most cases subsequently edited; however, in several instances the speakers opted to provide a formal paper. The statements made in these proceedings are those of the individual authors and do not necessarily represent the positions of the steering committee or the National Academies.

---

[1]For additional information on the "Workshop on Science and the Information Society," as well as ICSU's Declaration of Principles and the Agenda for Action, see http://www.icsu.org/.

[2]See, for example, National Research Council (NRC). 1997. *Bits of Power: Issues in Global Access to Scientific Data,* National Academy Press, Washington, D.C.; NRC. 1999. *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases,* National Academy Press, Washington, D.C.; NRC. 2000. *The Digital Dilemma: Intellectual Property Rights in the Information Age,* National Academy Press, Washington, D.C.; and NRC. 2002. *Resolving Conflicts Arising from the Privatization of Environmental Data,* National Academy Press, Washington, D.C.

[3]NRC. 2003. *The Role of Scientific and Technical Data and Information in the Public Domain,* National Academies Press, Washington, D.C.

Acknowledgments

The U.S. National Committee for CODATA and the Board on International Scientific Organizations of the National Research Council of the National Academies wish to express their sincere thanks to the many individuals who played significant roles in planning the International Symposium on Open Access and the Public Domain in Digital Data and Information for Science. The Symposium Planning Committee was chaired by M. G. K. Menon of LEAD India. Additional members of the Steering Committee were Carlos Correa, University of Buenos Aires, Argentina; Dialo Diop, Université Cheikh Anta Diop, Senegal; Farouk El-Baz, Boston University, United States; Dominique Foray, Centre National de la Recherche Scientifique, France; Alexei Gvishiani, Russian Academy of Sciences, Russia; Elizabeth Longworth, Industry New Zealand, New Zealand; Lulama Makhubela, South African National Research Foundation, South Africa; Erik Sandewall, Linköping University, Sweden; Mary Waltham, Publishing Consultant, United States; and Ferris Webster, University of Delaware, United States.

We also would like to thank the following individuals (in order of appearance) who made presentations during the workshop (see Appendix A for symposium agenda): Shuichi Iwata, University of Tokyo, Japan; M. G. K. Menon; David Dickson, SciDev.Net, United Kingdom; Elizabeth Longworth; Thomas Dreier, University of Karlsruhe, Germany; Alan Story, University of Kent Law School, United Kingdom; Robin Cowan, MERIT/ University of Maastricht, Netherlands; Koïchiro Matsuura, Director-General, UNESCO; Clemente Forero-Pineda, University of Bogota, Columbia; Chrisanthi Avgerou, London School of Economics, United Kingdom; Massey Beveridge, University of Toronto, Canada; Jean Luc Poncelet, Pan American Health Organization, United States; D. K. Sahu, JPM Managing Editor, India; Leslie Chan, Bioline, Canada; Farouk El-Baz; Mukund Rao, Indian Space Research Organisation, India; Peter Weiss, U.S. National Weather Service, United States; Liu Chuang, Chinese Academy of Sciences, China; Andrew Kaniki, National Research Foundation, South Africa; R. Stephen Berry, University of Chicago, United States; Mikhail Zgurovsky, National Technical University of Ukraine, Ukraine; Jerome Reichman, Duke University Law School, United States; Charles Schweik, University of Massachusetts, United States; Erik Sandewall; Harlan Onsrud, University of Maine, United States; Sarah Durrant, International Network for the Availability of Scientific Publications, United Kingdom; Gilberto Câmara, National Institute for Space Research, Brazil; Saloshini Muthayan, Doctoral Candidate, South Africa; Florence Muinde, UNESCO Fellow, Kenya; Ndaendelao (Emma) Noongo and Nico Willemse, Ministry of Environment and Tourism, Namibia; and T. B. Rajashekar, National Centre for Science Information, India.

This volume has been reviewed in draft form by individuals chosen for their technical expertise in accordance with procedures approved by the National Academies' Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for quality. The review comments and draft manuscript remain confidential to protect the integrity of the process.

We wish to thank the following individuals for their review of selected papers: William Anderson, Praxis101; Carlos Correa, University of Buenos Aires; Paul David, Stanford University; Jean Garnier, National Institutes of Health; Bernt Hugenholtz, University of Amsterdam; Heather Joseph, BioOne; Stephen Rossouw, University of Capetown; and Gordon Wood, Canada Institute for Scientific and Technical Information.

Although the reviewers listed above have provided constructive comments and suggestions, they were not asked to endorse the content of the individual papers. Responsibility for the final content of the papers rests with the individual authors.

The U.S. National Committee for CODATA would like to recognize the contributions of the following National Research Council staff and consultants. Paul Uhlir, director of International Scientific and Technical Information Programs, was project director of the symposium. Julie Esanu helped to organize the symposium and served as the primary editor of the proceedings. Valerie Theberge organized and coordinated the logistical arrangements, and Amy Franklin assisted with the production of the manuscript.

In addition, the committee would like to thank the other individuals who contributed to the success of the workshop. Kathleen Cass, executive director of CODATA; Carthage Smith, assistant director of ICSU; John Rose, senior program specialist of the Information Society Division of UNESCO; and Barry Mahon, executive director of ICSTI, were integral in organizing the workshop. Malene Munkebo, communication coordinator of CODATA, and Annick Ongouya, a secretary at UNESCO, provided local logistical support in Paris.

Contents

Session 5: Innovative Models for Public-Domain Production of and Open Access to Scientific and Technical Data and Information

Session 6: Examples of New Initiatives in Developing Countries

Appendixes

INTRODUCTION

1

Welcome by CODATA President

Shuichi Iwata
University of Tokyo, Japan

I would like to express my gratitude and extend my heartiest welcome to all participants who have come from all over the world to attend this international symposium on open access and the public domain in digital data and information for science. This symposium is jointly organized by the International Council for Science (ICSU), the United Nations Educational, Scientific and Cultural Organization (UNESCO), the U.S. National Academies, the Committee on Data for Science and Technology (CODATA), and the International Council for Scientific and Technical Information (ICSTI). The John D. and Catherine T. MacArthur Foundation and the U.S. National Weather Service also provided support for this workshop.

The United Nations World Summit on the Information Society (WSIS) takes place in Geneva in December 2003 and in Tunis in 2005. The scientific and technological community has a key role to play in the development of an information society for the benefit of society as a whole. The overall aim of this symposium is to identify and define the key messages that the international scientific and technical community wants to convey to the other stakeholders in the information society during the WSIS process.

Data, information, and knowledge prepared by scientific and technological communities, in principle, are public goods, and they should be shared by everyone. If a set of data, information, or knowledge is simple enough for everyone to understand its meaning there are no problems in following this public-good principle. However, most of the data, information, and knowledge are complex and are usually processed, which adds different meanings, values, and sometimes barriers.

Even if these data become available in the public domain, many technical and societal issues arise. Who can take advantage of the vast amount of scientific and technical data, information, and knowledge and who cannot? How can we deal with terabyte data of different semantics, qualities, and sometimes with intellectual property rights?

CODATA'S ROLE IN PROMOTING OPEN ACCESS

CODATA has recently undertaken several activities that address these issues. Together with ICSU, CODATA has sponsored an ad hoc Group on Data and Information since 1997, which has focused on the importance of full and open access to scientific data and information on a global basis. This group submitted a white paper on these issues to the World Intellectual Property Organization in 1997 and subsequently developed a set of guiding

principles.[1] The ICSU/CODATA ad hoc group held a workshop in Baveno, Italy, on the "European Union's Directive on the Legal Protection of Databases." CODATA also held several sessions on these issues at its 2002 international conference in Montreal, Canada.[2]

In addition, ICSTI, working with ICSU and CODATA, held a meeting on the important topic of preservation of digital content, more particularly on the topic of continuous availability of digitally produced materials. There are significant technical, administrative, and economic issues associated with the longer-term availability of scientific data and information, and these constitute important elements in establishing policies in the area of open access.[3]

More recently CODATA has worked with ICSU and UNESCO in promoting the role of scientific information in the emerging information society, as input to the WSIS. CODATA is working with ICSU and UNESCO in planning a workshop on "Science and the Information Society," which will immediately follow this symposium.[4] This workshop will bring together scientific experts, managers, and representatives from several intergovernmental agencies to try and identify the major issues for science in relation to the WSIS. The product of this workshop will be an agenda for action that will be submitted to the WSIS.

---

[1]See *Access to Databases: Principles for Science in the Internet Era* at http://www.codata.org/codata/data_access/principles.html.

[2]See the keynote and related plenary session on "Legal Issues in Using and Sharing Scientific and Technical Data," as well as the plenary session on "Information Economics for Scientific and Technical Data," in the 2002 CODATA Conference Proceedings, which can be found at http://www.codata.org/codata02/index.html/.

[3]ICSTI, along with INSERM and INIST, also organized an international meeting on January 23-24, 2003, on the subject of "Open Access to STI: State of the Art and Future Trends"; see the INIST Web site at http://www.inist.fr/openaccess/index_en.php for additional information. This meeting evaluated the consequences of recent organizational and technical developments on the wider availability of scientific information. The meeting also dealt with the issues from the point of view of authors, publishers, and audiences, especially those in the developing world, and policy makers.

[4]For additional information on the "Workshop on Science and the Information Society," see http://www.icsu.org.

2

Introduction by Symposium Chair

M. G. K. Menon
LEAD, India

I am honored to chair this meeting and to participate in an area of crucial significance for science and human development. This symposium will focus on issues relating to scientific data and information and on access to these resources, particularly from the viewpoint of those created by public funding, which should be in the public domain and on which there are significant encroachments taking place.

Science today not only produces data but also is significantly dependent on these data for its own growth and survival. Science is highly interdisciplinary. One cannot say at any point in time, in terms of data or information, what may be required. To work in biology and genomics research one needs access to data from functional genomics, structural biology, and crystallography, apart from much else. The same data from crystallography may be needed for research in physics or chemistry. The various disciplines within science are linked by their need for access to the same databases. Data are fundamental to the growth of science. Any effort to restrict access to data will, in some way, prevent the growth of science.

Science has been based on the availability of information, on open access to that information, and on transparency. It results in further growth of science as well as technology and applications. It is in the area of technology and applications that intellectual property rights should apply. But the point I wish to emphasize is that the primary source of information and data cannot be denied to science.

Databases are becoming enormous and interlinked. For example, consider the global earth system and its many interconnections. The International Geosphere-Biosphere Program of ICSU is a very large program. In asking a question on temperature rise and its consequences, one might first consider carbon dioxide and then other greenhouse gases in the atmosphere. One then looks at the generation of these gases and at the way they are transported. The gases may remain as they are or be absorbed, converted, or stored in repositories. This then leads into areas like oceanography, including sea-level rise, and plant systems. A range of such issues arise, which is why ICSU included the biosphere program in the International Geosphere-Biosphere Program. Similarly, if you consider biodiversity and what exists on Earth, it is large and complex. One would like to know more about relationships and various aspects of evolution. People ask the question whether all of these data can be put together in databases and accessed; this is a separate question for computer scientists. On the other hand, when one looks at the scale and complexity of databases, and wants to access pieces of it, then one is often faced with intellectual property restrictions.

A further feature of databases, which very often is not considered, is their invisibility. In a sense they are inherent in many different places and get absorbed into areas of application and user interfaces. In these cases we have to be very careful about restrictions that apply.

In addition to its responsibility in the production of data and databases across many disciplines, science has also provided the underpinning of what we now refer to as the information and communication technology (ICT) revolution. Physics and chemistry were basic to the advance in solid-state electronics and microelectronics. Mathematics was fundamental to the development of software. It was, in fact, the vast amount of data pouring out from the high-energy particle accelerators that created the conditions for the development of the World Wide Web, which we all use. The World Wide Web originated at CERN in Geneva.

Science created the Internet. Many of the tools of information technology—for example, space and communication technologies, including knowledge of the radio spectrum, the laser, and aspects of magnetism needed for storage—all emerged from it. Science has made it possible, through ICT, to have databases and access to them. We cannot deny science, which has been the creator of all of this, its future. We must see to it that access is widely available.

CHALLENGES FOR DEVELOPING COUNTRIES

We also must consider the impact of restrictions on access, not only from the viewpoint of science and as an international endeavor, particularly noting what is happening in the developed countries, but also on what would happen to the developing societies, and how they would access the data and information that is vital for their development and for their nascent science. We are talking about an economic divide, and now there is the growing digital divide. The problems faced in developing societies, including poverty, illiteracy, population growth, malnutrition, poor sanitation, and ill health, among others, are a result of poor economic growth. This will be exacerbated by the growing digital divide.

First of all, the poor cannot afford computers. Telecommunications in general in developing societies is poor and where (and to the extent) available, expensive, making access to the Internet limited.

Illiteracy is especially a problem. People who are illiterate cannot use computers. The Internet is significantly English-dominated and will be so for a long time to come. In India today only 5 percent of its population of one billion know English. Although there is access in some other major languages, none of these are the primary mother tongues or languages in developing societies. There are vast numbers in the developing societies who, for a variety of reasons, are educated but computer illiterate. In most places energy availability is poor and in some cases totally unreliable for the use of current computer systems. Computer systems will have to be made more inexpensive and telecommunications less costly for more widespread growth of information technology in these countries.

These are the problems developing countries face, even if total access to everything is available, without any hurdles and barriers relating to access. Reading material has become enormously expensive. There are very few universities in developing societies that can afford the scientific journals that are produced. These journals are enormously expensive, along with books; therefore, as far as developing societies are concerned, one has to consider the ethical aspects of this issue. If you want them to develop, it is absolutely vital that they be given freer, much more open and lower-cost access to the existing and growing global reservoir of global knowledge. In return they have an enormous valuable inheritance in content that they can provide to the world, so far carried by oral tradition, community knowledge, and social interactions. Its magnitude and value have still to be evaluated, but it is certainly significant.

3

UNESCO's Approach to Open-Access and Public-Domain Information

Koïchiro Matsuura
UNESCO, France

The rapid evolution of knowledge societies continues to provide new means for achieving progress in all sectors of work and life through the increasing use of information and communication technologies (ICTs) such as computers and networks. While ICTs have greatly facilitated the movement and handling of data, the process of generating and validating information and knowledge remains essentially one of human creativity. While access to the information highways is still a real problem in many countries, questions of access to scientific and development data and information in the digital world, including questions of intellectual property rights, are attracting growing debate.

Science and education are at the very center of debates on the challenges and opportunities of knowledge societies. We face a paradox, however. On the one hand the accelerating spread of the Internet and new opportunities for free or low-cost publishing are generating real benefits. On the other hand, the new economic and technological environment is raising concerns about the erosion of access to certain information and knowledge whose free sharing facilitated scientific research and education in past decades.

Before proceeding to outline some aspects of the UNESCO approach to addressing these concerns, I would like to stress two fundamentals that govern our action. First, in UNESCO's view the concept of "knowledge societies" is preferable to that of "the information society" because it better captures the complexity and dynamism of the changes taking place. The knowledge in question is important not only for economic growth but also for empowering and developing all parts of society. Thus, the role of new ICTs extends to human development more generally, and therefore to such matters as intellectual cooperation, lifelong learning, and basic human values and rights.

Second, most developing countries have thus far been unable to take full advantage of the advances offered by new ICTs in terms of access to scientific and technological information and learning opportunities, at least relative to the situation in the industrialized countries—the "digital divide." If knowledge societies capable of generating new knowledge in a cumulative, cooperative, and inclusive process are to be created, they should be based on shared principles, particularly equitable access to education and knowledge. National policies, supported by international frameworks, can be a tool to facilitate access for all to essential information.

A key component of such frameworks and policies is the work of the United Nations system, under the leadership of the World Intellectual Property Organization (WIPO), to develop balanced and consistent international standards for copyright and neighboring rights as exemplified in the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty, both adopted in 1996. UNESCO's policy is to encourage and assist

member states to promote access to information and knowledge for the progress of science and the diffusion of education, keeping in mind the necessity of rigorous conformity with international conventions on intellectual property.

It is in this spirit that UNESCO has been working, following a directive of its General Conference in 1997, to elaborate a draft recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace.[1] When submitted to the General Conference at its last session two years ago, no consensus was forthcoming. I therefore intervened personally by proposing that further structured consultation involving member states, other stakeholders, and experts be organized by the Secretariat with a view to generating greater consensus.

This process of consensus building has moved forward successfully and the revised draft recommendation will be considered by the General Conference at its 32nd session, where it likely will be adopted. It will then be presented as a contribution of UNESCO to the first World Summit on the Information Society to be held in Geneva in December 2003. Two of the four main sections of the proposed draft recommendation—"development of public domain content" and "reaffirming the equitable balance between the interests of rights-holders and the public interest"—bear directly on the themes of this symposium. Without attempting to be comprehensive in the short time available, I will allude to some of the major points of the proposed draft recommendation in discussing the action that UNESCO is undertaking to promote practical solutions in these areas.

Public-domain information is publicly accessible information, the use of which does not infringe any legal right or any obligation of confidentiality. It thus refers to the realm of all works or objects of related rights that can be exploited by everybody without any authorization. While many people associate the public domain mainly with classical and traditional literature, an equally important store of public-domain information for development, and undoubtedly most important for science, is public data and official information produced and voluntarily made available by governments or international organizations.

Another paradox arises here. Public-domain information, which is free of copyright, is often not sufficiently well known to potential contributors and users, and in some countries there are growing restrictions on the availability and use of public information and data. Such restrictions arise, for example, when information and data that are in the public domain become privatized or commercialized through a process of re-packaging.

In fact, the electronic public domain forms an international virtual public library that is vast and growing. This electronic public domain, furthermore, is both a world heritage and an invaluable support for productive commercial- and creative-sector activities in developing and industrialized countries. All would gain if governments and other public service organizations would identify and digitize their rich and diverse information stocks and make them available through the Internet. Thus, the UNESCO draft recommendation encourages member states to "recognize and enact the right of universal online access to public and government-held records" and to "identify and promote repositories of information and knowledge in the public domain and make them available to all."

UNESCO is encouraging this process in international forums and also in its advice to member states, notably through the preparation of "Policy Guidelines for the Development and Promotion of Public Domain Information."[2]

OPEN-ACCESS AND VOLUNTARY AUTHORIZATIONS

The public-domain principle can be extended conceptually by the assimilation of open-access information made freely available by its rights holders without cost. One well-known example of open access is the open-source software license by which computer programs are distributed free of charge by their authors for exploitation and cooperative development. Another is the vast amount of documentation produced and made available free of charge by the United Nations and its specialized agencies. Yet another is the movement of educational institutions around the world to provide their educational resources on the Internet free of charge for noncommercial usage,

[1] See http://portal.unesco.org/es/ev.php@URL_ID=1276&URL_DO=DO_TOPIC&URL_SECTION=201.html.
[2] See http://unesdoc.unesco.org/images/0012/001297/129725e.pdf.

typified by the OpenCourseWare project of the Massachusetts Institute of Technology. The UNESCO draft recommendation urges member states and international organizations to encourage open-access solutions, and UNESCO itself is strongly committed to promoting information sharing in education, the sciences, and culture, and to disseminating information and software for development under open-access conditions.

The question of access to commercially published information, which is of great importance to science, is a different one. It is to be noted that many publishers are interested in providing their works electronically under preferential conditions for science and education, particularly to users in developing countries, provided their copyright is strictly respected. Numerous international programs are now showing that affordable access to commercial publications in developing countries is possible: notably the Health InterNetwork Access to Research Initiative of the World Health Organization; the Pharmaceutical Education and Research Institute initiative of the International Network for the Availability of Scientific Publications, created by ICSU and UNESCO in 1991; the electronic-Journals Delivery System of the Abdus Salam International Centre for Theoretical Physics in Trieste, Italy; and the more recent, ambitious initiative of the revived Alexandria Library, from its beginning a UNESCO project, to make virtually all of the world's books available through the Internet. UNESCO is looking carefully at ways to promote this type of initiative, for example, through model frameworks of voluntary permissions by which publishers and other rights holders could assign specific rights to users in developing countries, either definitively or on a limited time basis.

Another quite distinct matter concerns provisions for a fair balance of interests in the use of copyrighted works in the digital environment. This refers to the limitations and exceptions to copyright and related rights protection that are authorized in national legislation—as required in the two aforementioned WIPO treaties—provided that they are applied only in certain special cases that do not conflict with normal exploitation of the work and do not unreasonably prejudice the legitimate interests of rights holders. Such provisions for equitable use in the public interest, which vary from country to country and are sometimes called "fair use," "fair dealing," or "limitations and exceptions authorized by the law" in specific legal systems, typically provide for exceptional free reproduction of copyrighted information for such uses as education, research, library services, journalism, and access for disabled persons. These equitable use provisions, which in the predigital world made possible the public library, are potentially of even greater importance in the digital world. However, they also present greater risks to the legitimate interests of rights holders given the ease with which digital information can be redistributed once released.

UNESCO recognizes the importance of equitable use provisions in national policies in education, the sciences, and culture, particularly for the developing countries. We also recognize the importance of a fair balance between the interests of rights holders and those of users when cultural works and performances are exploited in the digital environment in the fields of teaching, scientific research, libraries, dissemination of information, and the needs of the visually impaired. In this regard UNESCO, in close consultation with the concerned user and rights-holder communities, is carrying out an extensive study aimed at comparing the relevant provisions in existing national legislation with actual needs. Later on, a consensus-building process will be proposed regarding how best to address in practice any identified gaps, paying full respect to relevant provisions in WIPO and World Trade Organization treaties and without undermining copyright protection.

4

Science Communication and Public Policy

David Dickson
Science and Development Network, United Kingdom

INTRODUCTION

In this presentation I will try to place the issues that will be addressed at this symposium in a slightly broader context than they are frequently seen. The key issue we face is, who should have control over access to scientific knowledge and the terms on which that knowledge is used? One of the core political challenges facing the world today is the threat to the public domain usage of that knowledge coming from various directions, including parts of the private sector.

The central role of science in providing the base of the technology that we use in virtually all aspects of our daily lives, and which applies in both developed and developing countries, raises a question that concerns us all. Who will determine the shape of the future?

That is a rather broad and grandiose question when put it in those terms. I want to narrow it down slightly to address the more specific issue of public policy on science-related issues, including topics ranging from genetically modified crops to climate change to research on tropical diseases. I will focus in particular on the impact of science communication, a term that I use to describe the way information about science is communicated to the public, and not just within the research community.

SCIENCE, COMMUNICATION, AND PUBLIC POLICY

This argument has three main components. The first is that the way the policy on science-related issues has been going in recent years has taken a significant shift. Until recently science advice was essentially a linear affair. Governments sought advice from a relatively restricted group of advisers drawn mainly from the scientific community. This advice, once it had become integrated into the political priorities of the politicians involved, formed the basis for much of science policy, or policy about science.

More recently, however, sensitivity to growing public concern over the impacts of science on society has meant that the advice of scientists is increasingly seen as only one among a number of inputs into the policy making process. The voices of other stakeholders, which can range from environmental groups to religious organizations expressed either directly or through the mass media, have come to play an equally important role in shaping the judgment of politicians and government decision makers. Consider, for example, in recent debates

over human cloning and embryo research or genetically modified crops the extent to which public perceptions are molding government policy.

As a result science policy is no longer this linear affair but one that is based on networks of information, networks in which science communication has a vital role to play, and in which the Internet and electronic communication have come to play an important part.

The second component of this argument, therefore, is that the development of electronic communication, the Internet, and the World Wide Web have stimulated and reinforced this process. One of the strengths of the Internet is its ability to democratize access to information. We have seen this in Eastern Europe, where it helped to point the way to the downfall of political systems whose previous strength had been partly in their ability to control access to information. We also see it in the field of science policy where, at least in principle, every citizen now has the possibility of obtaining direct access to a wealth of information about the science and technology that affect their lives.

However—and this is the third component of the argument—we should not read too much into the power of the Internet. For even if it has encouraged this shift in the nature of decision making on science-related issues, it has not by itself changed the nature of the forces that control access to that information. It was people, not the Internet, who brought down the Communist states of Eastern Europe. And, at a more prosaic level, it is social institutions, not electronic technology, that control access to scientific data in the digital era.

Indeed the current debate over open access to scientific information in many ways is merely the contemporary manifestation of issues that have been around for many years. I wrote a book about 15 years ago on the nature of science and science politics in America, which had very much the same thesis, namely that we were witnessing at that point a pre-Internet battle over access to and control over scientific information.[1]

To summarize, my thesis has three components. First, policy making on science-related issues is going through a dramatic shift from linear information flow to networks. Second, electronic communication through the Internet is promoting and supporting this shift. But third, questions of control over and access to scientific information remain very much as they were previously, even if the form in which these issues manifested themselves has been significantly transformed.

THE ROLE OF SCIENCE COMMUNICATION IN SOCIETY

To help us focus on the issue of science communication I would like to refer to one of the most significant images of the science and society debate over the last two decades. It is a picture of Agricultural minister John Gummer in May 1990 feeding a hamburger to his somewhat bemused and reluctant daughter, Cordelia. The typical caption that accompanied the picture usually read, "UK Agricultural Minister declares that there was no need to worry, British beef is safe." The accompanying news story quoted Gummer as saying something along the lines of "British cows may be dying of mad cow disease, but I am so confident that it is not a threat to human health that I, a responsible parent, am personally prepared to stuff British beef into my daughter's mouth."

The rest, as they say, is history. It was not long before Gummer and indeed the whole British government had to eat their words—almost literally—and admit that they were wrong. Bovine spongiform encephalopathy (BSE) can pass into the food chain with tragic consequences. Furthermore, this particular picture has come to haunt Gummer—who ironically had a relatively good record as a defender of the environment—as well as the Conservative Party and government public relations officers ever since.

What do we make of this picture both then and now? It is essentially a political picture that attempts to convey a claimed scientific certainty, namely, that British beef is safe. With the benefit of hindsight the manipulation is obvious. We are now well aware of the real function of this image, the message it was intended to convey, and the claim to scientific legitimacy on which it was intended to be based. Indeed, the subsequent realization by the British public of this particular manipulation and of the whole government handling of the BSE debacle is widely blamed for a significant drop in the public's trust of both politicians and the scientists who advised them; and in the

[1]David Dickson. 1986. *The New Politics of Science,* University of Chicago Press, Chicago.

process it is part of what has contributed to this shift in the nature of the science advisory process. The trust has been broken in the linear system, and has helped transform the process into a network scheme.

The use of the image in this way also presents a challenge. How do we expect readers to relate to it? In particular, how do we as journalists balance our desire to inform both ourselves and our relative constituencies about the scientific perspective on critical issues such as the preceding, with the skepticism that is or should be a central component of our professional responsibility when it comes to interpreting statements that are intended to carry the authority of scientific truth?

The issues raised by the BSE crisis in Britain illustrate the critical role played by the media in articulating the relationship between science and society. This role, and the responsibilities that go with it, concern the accurate transmission of information.

I am referring to something that is much more than what is often described, somewhat disparagingly, as the "deficit model" of the public understanding of science, namely, the idea that the public lacks adequate knowledge of the facts produced by science. It includes, first, an equal responsibility for journalists to report on the uncertainties of such knowledge. It also means that journalists have a responsibility to report equally accurately on the impacts of science and society, and the responses to such an impact.

All these things are relatively conventional knowledge within the journalistic community. Journalists are also keenly aware that achieving a proper public understanding of science is actually a two-way process that must include the scientists' better understanding of the public. This concept is just becoming recognized within the scientific community itself. To quote the words of Alan Leshner, executive officer of the American Association for the Advancement of Science, in a recent issue of *Science*: "Simply trying to educate the public about specific science-based issues is not working in this one-way information flow. Given the uncertainty in science, the best science-based strategy is not always as clear as we would like and as many in our community might claim."[2] He goes on to say that we should move beyond paternalism. We should engage the public in a more open and honest bidirectional dialogue about science, technology, and their products, including not only their benefits but also their limits. We should respect the public's perspectives and concerns even when we do not fully share them, and we should develop a partnership. Indeed I would argue that we should go beyond these sentiments to ask how this process of developing partnership with the public is likely to work out in practice and who will be engaged in establishing it.

One system that works highly effectively in this arena is the scientific press, the area of science communicators. To see how this operates it is important to understand the extent to which the role of a journalist is not one of simply conveying the truth to the public. Journalists are not conveyor belts who report on facts any more than the role of a scientist can be defined as simply discovering scientific facts. Rather a journalist's role is to report on significant facts and, if space allows, on the nature of the significance. "Dog bites man is not a story," but "Man bites dog" certainly is. This is because it is so unusual that we want to know when, who, why, and how—and did the dog bite back!

In other words the task of any journalist—and science journalists are no exception—is essentially one of extracting significance from a mass of evidence, policy documents, and headline-grabbing statements from individuals and institutions who may or may not have a vested interest in their outcome. To that extent there is an important role in which the science journalist, or the science communicator, acts as a proxy for the public when it comes to interpreting and articulating the relationship between science and society or, to put it another way, between knowledge and power.

The terms "interpreting" and "articulating" are both somewhat abstract concepts. What they are intended to convey is a sense that the way the media handles science has essentially become an important constitutive component of the policy-making process on science-related issues. The media does more than just report policy choices to the public on such issues, or even report on the response of the public to the policy choices. The media also helps to frame these policy issues and the public responses to them, and this is why full access to information is so important.

[2]Alan Leshner. 2003. "Public Engagement with Science," *Science* 299 (14 February).

THE SCIENCE AND DEVELOPMENT NETWORK

The Science and Development Network (SciDev.Net)[3] is attempting to promote such goals within the developing world and show how the Internet can assist in improving access to scientific and technical information for developing countries. SciDev.Net seeks to empower people at all levels of these societies by providing access to authoritative information and informed opinion on key issues at the interface between science and society, part of the networks described previously.

At the same time, we are attempting to promote informed debate on such issues by enhancing the professional skills and increasing an awareness of professional responsibilities of all those engaged in a process that can be called science communication. The goal of this project is to empower all members of civil society in the developing world to make more informed choices on the critical decisions they face in pursuit of sustainable development.

SciDev.Net was set up at the end of 2001. It is backed by the journals *Nature* and *Science* with additional support from the Third World Academy of Sciences. The form of this support from *Nature* and *Science* is actually very important: both journals have provided free access to a select number of articles from each issue every week. Access to these articles is actually by subscription only, but as a concession they have agreed to provide privileged free access to selected articles. People can access these papers, which are relevant to the needs of developing countries, without cost on the Web site.

SciDev.Net has three main activities. First, we operate as a free access news and policy-oriented Web site that includes dossiers containing policy briefs and news and opinion articles on key topics of current interest; the news will occasionally include complete access to original scientific papers. Second, we are building regional networks of individuals and institutions that are committed to improved science and technology communication. Third, we organize workshops and other meetings and do capacity building in this field.

The Web site has news, feature opinion articles, and weekly editorials, which are relatively conventional journalistic items.

In addition, however, the Web site has two features made possible by the electronic revolution that are especially important. First, we include the dossiers, which enable us to compile current and previous information on particular topics. The second is the set of links to other scientific organizations and journals.

SciDev.Net also tries to help build capacity in science communication within the developing world in various ways. For example, the site includes information about writing a scientific paper or submitting a paper to a scientific journal. We try to get experts in the field, when possible from the developing countries, to provide this basic information.

SciDev.Net launched, as previously stated, officially in December 2001. We have been growing at a steady rate largely by word of mouth. We propagate, advertise, and communicate with people through the Internet. That is another example of how the Internet is very important to our activities.

SciDev.Net has regional networks in Africa and Latin America. There are plans to launch a network in South Asia, which will probably be based in India, toward the end of 2003. We are also talking to the Chinese Academy of Sciences about establishing a network covering East Asia, which will likely be based in Beijing. SciDev.Net is very keen to cover the Middle East and North Africa, and we also cover Southeast Asia.

As previously mentioned, SciDev.Net also convenes workshops. For example, we held a workshop on sustainable development during the World Summit in Johannesburg in 2002. Another workshop was held in Entebbe, Uganda, on science, communication, and sustainable development. This was for journalists from eastern and southern Africa, and was held in October 2002.

This raises a very important issue, as Professor Menon mentioned,[4] of access by developing countries to this type of technology. It is very low at the moment, but the demand and the uses are there. SciDev.Net does not pretend to get its message across directly to the man or woman in the street, so to speak. We are getting our message to the communicators, to the people who are in contact with the people in the street, and those people do use the Internet.

---

[3]See the Science and Development Network at http://www.scidev.net.

[4]See Chapter 2 of these *Proceedings*, "Introduction by Symposium Chair," by M. G. K. Menon.

In conclusion, SciDev.Net is trying to develop a new way of providing the public with access to information about science that on the one hand makes use of the exciting technological possibilities being opened by the Internet, but on the other hand uses very traditional journalistic and communication devices. We include peer review and editorial control. We maintain a clear editorial identity in what we are doing because we also believe that that is part of the essence of getting this communication across. We hope that this combination provides a valuable tool for helping to increase access to information about science and technology in the developing world.

SESSION 1: LEGAL, ECONOMIC, AND TECHNOLOGICAL FRAMEWORK FOR OPEN ACCESS AND THE PUBLIC DOMAIN IN DIGITAL DATA AND INFORMATION FOR SCIENCE

5

Introductory Remarks by Session Chair

Elizabeth Longworth
Industry New Zealand

This first session will consider the legal, economic, and technological framework for open access and the public domain, with a focus on scientific and digital information. There has been a huge shift in understanding the value of digital data and scientific information. It is now seen as absolutely critical in terms of economic development that there be a renewed focus on the importance of making this type of information and data openly available.

It is worthwhile to reflect on why this matters. The debate for some years has been framed around issues of the digital divide, concerns that if people cannot participate in accessing scientific and technical knowledge then they are severely debilitated in terms of growth, the progress of their community, social cohesion, and inclusiveness.

There is another reason why this is important. Considering the movement from agrarian societies to industrial societies to information societies to knowledge societies, Dr. Herman Hauser of Amadeus Projects at Cambridge University talks about the future as being "wet," and by "wet" he means biotechnology, genomics, and nanotechnology. As this "wet" future unfolds, the turmoil we have seen in ICT will be eclipsed by the increase in the importance of access to scientific and technical data and information.

Yet another reason why these discussions are so important is the huge shortage of investment in research and development, and in our scientific and education communities. A better understanding of the connections among scientific research, education, and the economic effects of freeing up access to data can help policy makers make informed investment decisions.

One issue in convergence is the difficulty we have had in talking about public domain and open access. We have not had the right language to describe what we mean. When we talk about the need to rebalance some of our existing innovation systems, it is often wrongly interpreted as an attack on intellectual property rights. We are challenged to fundamentally reexamine where the balance lies and look at the issues raised by that examination. This could be facilitated by clarifying certain definitions.

Information in the public domain means there is no exclusive intellectual property (IP) right in the information (e.g., the information is exempt from statutory IP protection, such as certain government information, or the statutory term of protection has lapsed). It also might be protectable subject matter, but by way of contract the information has been designated as unprotected, and it is made available to the public with no reservation of rights.

The other term is open access. When we use that term we assume that there are some proprietary rights attached to the information, but the rights holder allows that information to be made freely and openly available. There are tensions around intellectual property rights. We are starting to understand the public bargains in

copyright law. But we are also developing digital rights management technology that uses contracts and licensing and bypasses the public bargain inherent in public copyright laws.

The speakers in this session will also discuss economic models. There is a lot of discussion about the economic models that underpin open-source technologies; the concept of trust as an intangible, but one that now has value; and the economic pressure being exerted on the research community. Difficulties posed by the scarcity of resources for research institutions force them to commercialize ideas and spin them out in entrepreneurial activities. This leads to a real pressure for universities to distinguish themselves as centers of excellence, and the resulting competition puts enormous pressure on the historic ethos of openness.

Indigenous rights are also an issue area that often is overlooked. Communal rights are not accounted for by private rights or as intellectual property rights. But it is not appropriate simply to allow anyone to use communal and indigenous knowledge without consulting the knowledge holders.

There are also the issues with the technology platforms. While the Internet provides an infrastructure that supports intercommunication, we know that telecommunications companies are in crisis. The rollout of broadband mobile Internet services promises new business and pricing models. There are interoperability developments, including technical standards, metadata, and open networks. The good news is that these new economic models recognize collaborative ventures and the need for business and research institutions to work together to share information.

6

Overview of Legal Aspects in the European Union

Thomas Dreier
University of Karlsruhe, Germany

There is often much confusion about the workings and mechanisms of copyright protection. Copyright is a framework and not much more. Copyright can be used in order to favor access and it can also be used at the disadvantage of open access; it can have an enabling effect or a locked-in effect. However, copyright is often held responsible for blocking access.

COPYRIGHT AND OPEN ACCESS

Let us start with the idea of copyright and free access: are they friends or foes? We must consider some fundamentals of copyright. What is copyright? Copyright is an exclusive right. For nonlawyers one can imagine copyright along the same lines as a property right in tangibles. A property right is an absolute right over a good, which enables its holder to exclude others from using the particular good subject to the exclusive right and it gives its holder the right to use the good. Why do exclusive rights exist? Economists tell us that an immaterial, intangible good, such as data or information, has all the characteristics of public goods. Economists also tell us that without any incentive to take good care of public goods or create public goods there will be underinvestment in public goods. Hence, the legislature has decided to create an exclusive right, turning public goods into nonpublic goods.

What do exclusive rights enable right holders to do? Of course, an exclusive right allows its holder to exclude. However, in the field of copyright you generally do not want to exclude others arbitrarily, but rather you want to make sure that by allowing certain use and not others you are properly remunerated. The idea of getting money in return for the work you have created is an incentive to create. The author should be remunerated and the publisher, the producer who puts up the investment to disseminate the work, should recoup his investment. So, initially the idea of copyright is not to exclude others per se but only to exclude others in order to be able to grant an exclusive license to one particular producer. For example, authors may wish to grant the exclusive translation and publishing rights of their work to one particular producer or publisher and not to others, just to make sure that there is no illegal reprint of books, which would then transform the book as a product into a public good, which would deprive both authors and the producer of the original version of substantial income, which in turn would lead to an underinvestment into books.

The problem of exclusion of access may arise when intellectual property protection is granted to sole source products. A sole source product may be characterized as a product that has certain features that only come with this

one particular product. For example, we can all listen to music, which allows the user several choices. However, if you want to know the temperature in Paris yesterday between 9:30 a.m. and 9:45 a.m., that particular temperature might be stored in only one particular database. Possibly only one satellite recorded certain meteorological conditions a week ago over India and Bangladesh, which are found only in one particular source. If there is an exclusive right attached to such a sole source, you then have a locked-in effect.

The second problem that comes with the exclusive right is in the digital field. Producers are so afraid that their products will be copied, disseminated, and used without payment of adequate remuneration that they may feel tempted to block entire markets. Why is this a particular problem of the digital market? Well, digital goods can be repackaged, resold, and reformatted in many ways. As Jerry Reichman formulated: "Small incremental value may be added." All these incremental values may be quite useful to the end user, but initial producers tend to keep the market free of would-be competitors who start in product niches where incremental value is added to the initial product and who might quickly grow out of that niche into the way of the initial producer. But if the initial producer keeps the market niches free, this deprives us for the time being from having these useful value-added services. Needless to point out, these control strategies produce certain locked-in effects.

An example is publishers of newspapers and the electronic press-clipping services. For the time being, in many countries the publishers themselves do not offer electronic press-clipping services. Everybody would agree that it is highly desirable to have electronic press-clipping services. First of all, we could save a lot of paper and we would not have to chop down so many forests. Second, some trucks would not drive on our highways and pollute the air. It is much more environmentally friendly to have electronic press-clipping services. Yet publishers are very reluctant to allow third parties to come in because these third parties offering electronic press-clipping services on the basis of the primary publisher's product—namely the newspapers that contain these information services—apart from appropriating the market for digital exploitation of the initial product might even erode the basis for the primary product. Certainly we do not want our daily newspapers to disappear just because we gave too much freedom to third parties with electronic press-clipping services. Such services might be useful today, but if they then erode the basis of the primary product we do not have any value-added service anyway. But on the other hand, we want competition in markets of value-added services. To achieve this is a difficult task for copyright.

Third, let us briefly talk about free access and accessibility. The term "open" is very often confused not with mere accessibility but with "free" and this, moreover, sometimes in the sense of "cost-free" access. However, it is very clear that the production of information is cost intensive. Therefore, "free" in free access cannot mean free access as in free beer.

Fourth, even understood in the way just described, it should be noted that free access and accessibility are somehow linked to pricing. Users say, and the rights holders agree, that the price should be fair and reasonable. True, one has to pay a reasonable price in order to obtain access to certain information, but how high can the price be, or how low does the price have to be, so that we still can speak of open access or free access? It conveys dollar by dollar and if ultimately we pay $1,000 just to view one page, this very much looks like locking information away. Information would then no longer be free, no longer accessible. But what if the information is worth $1,000, or if it cost as much to produce the information? Who is to fix the price? In industrialized states, state authorities supervising prices were generally abolished after World War II, and the only legal instrument of at least indirectly controlling prices is antitrust law, which controls abuses of dominant market positions. But there is no such abuse as long as the right holder only makes use of his or her exclusive right. It follows that contrary to common opinion the distinction between free open access and information being locked away is a rather blurry one. Some energy should be devoted to clarifying these issues.

Another issue deals with access blocks by technological protection measures (TPM) and digital rights management (DRM) systems. It has often been said that the answer to the machine is in the machine. Initially TPMs were designed or understood as mere antipiracy devices, in essence, to ensure that works would not get copied. But there is more to it. Similarly, DRM has been understood as taking paper-negotiated contracts into the digital environment, but there too is more to it. The value added, so to speak, by TPMs and DRM is what economists call product diversification and price discrimination.

As an explanation, digital technology enables us to shape the use possibilities of one program in different ways. Let us take the example of music. If you receive a stream of musical data coming over the Internet, you are obtaining a listen-once-and-do-not-copy product. A CD, which is copy protected, allows you to listen as many times as you want but not to make copies. You could have another product further down the diversified product line that enables you to do more with that music, which would, for example, enable you to make one back up copy for use in your car. Ultimately, we still might have the "old-fashioned" CD as we know it today, which we could listen to and copy many times.

Again, economists tell us that with product diversification, producers can better capture the market. They can better answer to the market demand because there are some consumers who would be willing and able to pay more than they currently pay for the one-size-fits-all product, and there are others, who have not been able to pay the price for the current product, who could or might well be able to pay the (lower) price for a traditional CD. Product diversification by the user, and the ensuing price discrimination, raises the price for the high-use product and lowers the price for the low-use product. Economists tell us that this increases the overall public benefit. If TPMs are a good thing, we will have to make sure that consumers do not buy low-use products for a low price to circumvent the use-restricting TPM, and turn the product into a high-use product without paying. This, however, means that we need rather robust TPMs and strong legal protection against the circumvention of TPMs. This, of course, leads to the problem that effective technological protection measures can override the policy balancing that we find in our laws; again, we have a lock-in effect, this time unwanted by law.

Take the example of fair use. Fair use is permitted with regard to certain works, but if a work is protected by a TPM the system does not recognize whether the user wants to make commercial use of that particular work or just fair use as permitted by law. To this situation the law can respond in two ways. Either the law allows the user to circumvent the TPM in such a situation and thus risk that commercial use will be made without payment, or the legislature blocks the use completely, thus eliminating the freedom of fair use. In sum, introducing TPMs and DRM entails positive market effects, but it likewise results in rather strong lock-in effects.

THE LEGAL FRAMEWORK IN THE EUROPEAN UNION

Let us turn now to an overview of the legal framework in the European Union. First, a brief word on harmonization strategies in the field of copyright. The European Union has not created a community copyright similar to their community trademark, community design model, and community patent. Rather, an author obtains a bundle of national rights. These national rights are rather similar to each other so that there is no hindrance to the free movement of goods and services and no distortion of competition. It is the result of international conventions that makes this bundle seem as one unified right. An author creates a work in one country, and all the other countries that are parties to the international conventions promise to protect foreigners in the same way they protect their own nationals; they do so without asking for any formality requirements to be fulfilled. This makes copyright different from patents. For example, in order to get a British patent and a German patent you need two registrations. Of course, the process is simplified by the European Patent Office, where you can deposit one application, but the patentee still ends up with several national rights. With copyright you just create, and if you create in France you are treated in Germany, Great Britain, the Netherlands, and Belgium the same way as if a Belgian citizen had created a work in Belgium.

Within this bundle the copyrights are still distinct from each other. If there exists a limitation under U.K. law that declares certain use acts with regard to protected subject matter as fair dealing, then this only has its effects in Great Britain but not in Germany or France.

An interesting example is legal texts, which are free in Germany. There is some authority that even databases containing legal texts are free. In Germany I can download legislative material and make it available in my own database. However, the German author of legal texts could invoke his French copyright, since the acts I have performed in Germany, might also have—via the Internet—an effect in France and hence might violate the French copyright. Although the author of the material not protected in Germany cannot stop me from offering that

database for German Internet users, the author can still stop me for having that very same database made accessible to French users.

The purpose of copyright harmonization is to eliminate such discrepancies of national copyright laws. There is an E.U. directive on databases,[1] which has tremendous repercussions for the scientific output. The E.U. database directive creates a copyright that protects the originality of a database, the way the data are collected and assembled, which is a right that can be violated only if the whole structure of a database is copied. On top of this the database directive also introduced a *sui generis* right that merely protects the investment made into a database and which makes it illegal to copy and reuse substantial parts of a database. It is even illegal to use insubstantial parts of a database provided the use is undertaken in a systematic way and interferes with the normal exploitation of the database in question.

Another point very often overlooked is the 1998 E.U. directive on access control.[2] This started with a British case in which British tourists who wanted to watch encrypted BSkyB signals in Spain when they were on holiday purchased unauthorized decoders. BSkyB obtained legislation that made the production and sale of unauthorized decoders illegal. This ultimately led to the adoption of the European directive on access control, which protects any service that provides and allows access or is based on an access control, making it illegal to manufacture, import, or commercially advertise circumventing devices (it does not subject the act of circumventing itself to the prohibition).

In addition, the E.U. directive on copyright in the information society,[3] as regards technical protection measures, also subjects the very act of circumventing to liability. Anyone who circumvents TPMs without authorization is subject to cease-and-desist orders, damages, and fines. That is potentially very far reaching for two reasons. The first is that in such an economy what do we do with the fair use? Do we allow it and eventually open up the doors for illegal commercial uses or do we close the doors to fair use on the assumption that it is better to block some fair use in order to make sure that no commercial use is being made if there is a TPM applied to the subject matter, rather then allowing fair use with regard to digital TPM-controlled subject matter? The European Commission has decided that in the online field, for databases that are contractually made available, TPMs override the balanced approach of limitations and exceptions of copyright. The European Commission has always argued that this is justified when you contractually agree to what you pay for. In practice the decisive question then is, who can pay for what? If prices charged are too high, then legal remedies are unlikely to break this situation. Of course, if the maker of the database abuses a dominant market position, antitrust violation might be applicable. But first, antitrust only comes in late (i.e., it only applies if the abuse has already taken place) and second, antitrust law is not really about price control. There has to be very strong evidence of an abuse of the dominant market position. The mere fact that the price charged for access is much higher than the actual production cost does not in itself make it an antitrust case. The only solution might then be to let the market decide. In rare cases such as the one in which Scientology tried to stop critics from making extensive quotes from its founder's writings, the overriding effect of TPMs over statutorily granted exceptions, such as the citation right, might even result in a violation of the fundamental right of freedom of expression.

Another major problem with the European copyright information society directive is that it did not harmonize the limitations and exceptions. That is why I used that example with a database of legal text in Germany. While the database is perfectly free in Germany, it can be blocked in France by invoking French copyright. That is, of course, a big problem—we have in theory broad and far-reaching copyright harmonization, but it does not really harmonize copyright in Europe. The reason for this lack of harmonization is that member states do not like to give up

[1]See Directive 96/9/E.C. of the European Parliament and of the Council of March 11, 1996, on the legal protection of databases at http://europa.eu.int/ISPO/infosoc/legreg/docs/969ec.html.

[2]See Directive 98/84/E.C. of the European Parliament and of the Council of November 20, 1998, on the legal protection of services based on, or consisting of, conditional access, O.J.L. 320, November 28, 1998, at http://www.ebu.ch/departments/legal/pdf/leg_ref_ec_directive_conditional_access_201198.pdf.

[3]See Directive 2001/29/E.C. of the European Parliament and of the Council of May 22, 2001, on the harmonization of certain aspects of copyright and related rights in the information society, O.J.L. 167/10, June 22, 2001, at http://www.patent.gov.uk/copy/notices/pdf/implement.pdf.

their national legal traditions. Moreover, any change in shaping the limitations invariably affects whole market segments that are so fiercely being fought over. The prospects for statutory compromise solutions in this respect look indeed rather dim.

CONCLUSION

First, copyright is not necessarily hostile to open access. For example, the open-source philosophy in software uses copyright in order to keep access open. The idea is to grant a nonexclusive license to anybody upon the condition that whoever uses the open code grants back any value that person may have added and upon condition that these users do not commercially exploit that particular software. In other words, users of open-source software under a general public license grant back to the open-source community a use right in any alterations and additions made, and they do not "proprietize" it by including open-source code in proprietary software. So open source is based on copyright and uses copyright in order to make sure that those who cheat on the conditions might be excluded.

Second, copyright can be used both to secure access and to block accessibility. This applies also to TPM and DRM. The real problems result from a too broad legal protection mainly in the field of *sui generis* database protection. There are also the economic concerns regarding sole sourcers, global players, and those who really want to block markets that they have either no intention to serve themselves or where they want to seek monopoly rents. The open question then is, when the price adequately reflects consumer demand, what is a monopoly rent?

7

Database Protection in Countries of the South

Alan Story
University of Kent Law School, United Kingdom

INTRODUCTION

The spectrum of legal questions facing countries of the South that relate to this symposium's focus, open access to digital data and information for science, could be the topic for a conference in itself. This presentation will provide a general overview of the global protection of digital databases, with particular focus on these issues in the countries of the South, followed by an examination of balancing national interests in access to databases. However, taking advantage of the prerogatives granted to a speaker, I will begin by reviewing what this presentation will not cover.

First, I will not use the phrase "developing countries" because the words "developing countries" are in my view somewhat misleading by themselves. Many remain unconvinced that a number of countries are actually developing as opposed to becoming more integrated and more dominated by rich countries. What, for example, does "developing" actually mean? And analytic precision is lost if, for example, Brazil and Somalia are grouped together in a category called "developing countries." I want to instead use the phrase "countries of the South." I admit this is also problematic, and I add the forceful caveat that there is a great disparity between, among, and within such countries. For example, there are several "South Africas" and although located in southern Africa, South Africa differs greatly from other southern African countries.

Second, this presentation will not be a legal primer or explanation of the domestic laws covering databases in various countries of the South. Rather, the focus will be decidedly global and international and will examine particularly the database protections that such countries have become obliged to enforce within their own borders, namely intellectual property rights belonging to non-nationals who are foreign database owners situated in relatively rich developed northern countries, primarily those in Europe and the United States. To cite one example, all of South America and the Caribbean account for only 0.2 percent of all databases existing in the world in 2001. One of the principal questions I will ask is, what is the basis for this legal obligation of property protection, and the corollary question, does this approach work for their benefit?

Third, this paper will not discuss the myriad of bread-and-butter legal questions relating to contracts that may arise between, for example, the owner of a particular database and the users, in this case, researchers in a country of the South who may want to use that database. With databases, as with other forms of intellectual property, legal questions related to individual commercial contracts are taking on increasing importance. Indeed, often they are more important than strict intellectual property questions. This is not the forum to address such particulars of contracts and the unequal bargaining that frequently occurs between database owners and researchers in the South.

Fourth, this paper will not examine in any significant detail all the complex legal issues of database encryption and the prohibition on anticircumvention devices that Professor Dreier[1] mentioned and the threat that these new legal technological developments pose to database access. The new buzz phrase "digital rights management" is a particularly dangerous development for users of electronic information, including database users, and it essentially focuses on protection rather than management. It is being driven, as the European Association of Libraries explains, by technology and its limitations, and in the process may attempt to trump the proper and legal use of copyrighted materials.

At the same time, it should be recognized that the conditions under which southern scientists and researchers do their work, and hence the circumstances under which they acquire access to their databases, are often very different from those facing their colleagues in the North in much richer countries. I will look at this question in more detail when I examine the notion of national treatment of copyrighted works. Furthermore, a wide range of legal issues, including those involving intellectual property rights, both create and reinforce these differences.

Because of the nature of intellectual property and its duration as a property right stretching many years into the future (and, indeed, perhaps permanently under the terms of the European Union's database directive) intellectual property determines not only present but also future distributions and allocations of both information and wealth. A database right as a property right expresses a power relation between persons, in this case between the owners and the users of the database, and represents not only the state's grant of monopoly rights to a private party (the owners) but also that party's power over other people. In other words, the owner of intellectual property in databases can decide who conducts research and what they can research about.

Finally, and by way of information, the production and use of digital data necessarily implies access to computers themselves operated by some type of computer software. Here again, intellectual property issues should be appreciated as setting the context for public domain and access issues.

Under the terms of the agreement on Trade-Related Aspects of Intellectual Property, better known as the TRIPS agreement, computer software is protected as a literary work under the leading global international copyright regime, the Berne Convention. TRIPS and Berne presume a proprietary model of software and, in the case of the best known proprietary operating system, Microsoft Windows, users in the countries of the South must, as elsewhere, pay for an expensive licensing agreement to access this software. They must also abide by a series of legally and technically restrictive rules if they wish to access data and scientific information.

Other speakers will address the alternative nonproprietary approach, such as open-source or free software, and I will make but two short and related points. One, if they are interested in wider access to data and information issues, donor agencies and governments in rich countries of the North should cease privileging intellectual-property-protected proprietary software in their computer aid programs to poor countries and, for example, to educational institutions. Some government aid agencies, when pushed, even admit the advantages of nonproprietary systems. For example, in the United States under the U.S. Agency for International Development, there is the Leland Initiative, which ships PCs equipped with Microsoft products to poorer countries. The coordinator of this initiative has admitted that "on balance, we are for the cheapest and most affordable approach which would be open source." So, the very agency shipping these computers loaded with Microsoft Windows to Africa and Latin America admits that this is, in fact, not the best software system for those countries.

Second, we from the North are not encouraging nonproprietary formats. While conducting research last year for the U.K. Commission on Intellectual Property Rights about the reasons for the still limited (though rapidly expanding) use of free and open-source software in the countries of the South, I often heard this response: "If this open-source software is so good why are so many organizations, companies, governments, and rich countries still hooked on proprietary software?" (At my university, the University of Kent, only Microsoft Windows is available for users.) If UNESCO and other organizations want to give at least equal billing and equal status to nonproprietary formats, they should begin the process themselves by contributing to better and more equitable global access to information and by helping to break the reliance of poorer countries on proprietary software.

[1]See Chapter 6 of these *Proceedings*, "Overview of Legal Aspects in the European Union," by Thomas Dreier.

LEGAL OVERVIEW OF GLOBAL PROTECTION OF DIGITAL DATABASES

Moving now to the main focus of my remarks, the question of the legal protection of databases, at its most basic level, is relatively straightforward. Article 10.2 of the TRIPS agreement states that so-called "original databases" are considered as copyright-protected works in essentially the same fashion as books or photographs. I say "so-called original databases" because we should not forget that, on the one hand, "original" in copyright does not coincide with the dictionary definition of original. It does not mean new, novel, cutting edge, or innovative; it is a highly ideological and misleading word. On the other hand, the level of skill or creativity required to clear this originality hurdle is generally very low and variable in different jurisdictions. In the United Kingdom, for example, where I work and live, mundane work such as football pool coupons, street directories, and TV schedules have been protected as compilations equivalent to original databases.

Countries that wish to become members of the World Trade Organization (WTO) must also become signatories to the TRIPS agreement and hence provide copyright protection for all such databases. Indeed, such protection is automatically mandated among all signatories immediately upon creation of the database. In other words, if we today created an original database, it would automatically get copyright protection in France, and then all countries in the world that are TRIPS signatories would also have to recognize that it should be protected.

TRIPS, like the Berne Convention and the World Intellectual Property Organization (WIPO) Copyright Treaty, is a minimum rights agreement. A country must provide state protection within its boundaries up to a certain minimum level, which is in itself quite high, or it would be in breach of the agreement. For example, a country of the South is prevented from deciding whether to allow copyright protection in insect-related databases (original or nonoriginal) even though the country is strongly reliant on agriculture, has a very low income level, and is aware of the damage caused by insects and other pests and the low level of research on the problem. Nor could that country say that it does not want to have encryption devices in such databases. They could not make this decision, even if this provided a real incentive for the advancement of scientific research. Moreover, the TRIPS agreement does not allow broad public interest exemptions to copyright. For example, it is not possible to create an exemption based on an economic reliance on agriculture for sustainable food resources.

The protection of this copyrighted work, in this case a database, must last for a minimum number of years from the death of the author of that database. At the same time there is no maximum term of protection and it is quite legal for countries to establish a term of life plus 70 years or indeed to continue to extend the terms of such protection. This has already occurred in the United States and the European Union.

The unevenness of copyright database duration terms across the globe is necessarily trade distorting and contradicts the first clause of the first sentence of the preamble of the TRIPS agreement, which states, "Desiring to reduce distortions and impediments to international trade. . . ." In fact, this notion of allowing maximum terms while providing no ceiling itself distorts international trade. The overprotection of intellectual property rights is not, however, viewed as trade distorting.

The 1996 WIPO Copyright Treaty, currently ratified by 40 countries including countries of the South, also contains copyright protection for compilations of data or other material in any form, whether digital or not, and which, by reason of the originality of the selection or arrangement, constitute intellectual creations. This is the originality language also found in the TRIPS agreement, and it is meant to distinguish "original" databases from so-called "nonoriginal databases," which are not the subject of any international legal norms, but which are protected by countries of the European Union as well as Iceland, Liechtenstein, and Norway. Scientific information that exists outside the parameters of a database, for example, published in a journal, is also protected by copyright and this constitutes another barrier to access for scientists in the South.

THE BALANCING OF NATIONAL INTERESTS IN ACCESS TO DATABASES

As already stated, all WTO member countries that are necessarily signatories to TRIPS and all countries that are signatories to the WIPO Copyright Treaty must provide national treatment to all databases, both those produced within their own jurisdictions and those from all other WTO members. As Paul Goldstein explains, national treatment is a rule of nondiscrimination promising foreign traders who come within a treaty's protection that they

will enjoy the same treatment for their creations in the protecting country as the protecting country gives to its own nationals.[2] Or as Stephen Ladas, another commentator, puts it, national treatment means the complete assimilation of foreigners to nationals.[3] If WIPO succeeds in establishing international protection for nonoriginal databases, then we can expect to see the notion of national treatment in both original and nonoriginal databases, which countries of the South will also be expected to endorse and uphold.

At first this appears noncontroversial. Would we want to endorse discriminatory treatment of foreign databases within a particular national jurisdiction? If, for example, we were drafting new copyright or database legislation for a country such as Egypt, would we want to say that databases generated in Egypt should receive preferential treatment and protection within Egypt or, conversely, more or less open access compared with databases created in the United States or the European Union? Yet we should look into the matter more deeply. Should all databases, no matter what their national origin, be treated alike? Should an economically disadvantaged country be required to protect a database generated by a rich country as if it were generated within its own borders or within those of another poor country? More pointedly, does this "as if" approach really make sense? In other worlds, should all databases around the world be formally protected on the same level?

The answer suggested by other parts of the law suggests otherwise. Speaking about the conditions existing in Paris, Anatole France wrote in 1894 that "the law in its majestic equality forbids the rich as well as the poor to sleep under bridges, to beg in the streets, and to steal bread." A law whose very nature is based upon difference results not in equality but rather in substantial inequality, a difference in consequence and effect. The same is true of databases. While a database created in the United States and a database created in Namibia are both databases, just as a rich person and a poor person are both persons, the conditions that led to their creation are in all likelihood radically different. Why should they be treated alike when they are in fact very different? Why should there be a one-size-fits-all approach to databases, which requires the same conditions of access, the same cost of access for all researchers, and the same terms of protection? Is it equitable to require poor countries of the South to put their legal machinery in the service of the mission of treating "un-alikes" alike?

The other side of this "as if" problem suggests that the conditions of access should be the same in all countries. Professor Dreier talked about the question of fair use,[4] which would require that countries in the South, through protection of their databases, permit exemptions only in certain special cases. Such cases must not conflict with the normal exploitation of the work and must not unreasonably prejudice the legitimate interests of the rights holders, to use well-known phrases from international copyright conventions.

Let us examine these phrases. We cannot find out what these mean from the textbook writers and from the case law, but it is assumed that countries of the South in their approach to protection must track the notions of fair use in U.S. copyright jurisprudence, of fair dealing in the United Kingdom, or of similar exemptions available in Europe. The standard of what is considered "normal" is based on the standard existing in the United States, Europe, and a handful of rich countries, even though they have different traditions and different levels of income than countries of the South. So we can ask whether it is fair that what is considered fair and normal in France also becomes the standard for fairness in Senegal? Given the radically different income levels, why is it not fair that one set of researchers, for example, those working in France, should be charged one rate of access and another set of researchers, such as those in Senegal, be charged much less? International copyright law, however, does not ask these basic fairness questions.

No text, then, suggests that it would be fair or a legitimate special case for a country of the South to enact legislation that would recognize its special needs for access to data and information, even when such data and information have arisen after decades, sometimes centuries, of economic exploitation by rich countries and have led to global inequalities in educational attainment and literacy. It would not, for example, be allowable for South Africa to say that it was going to create an exemption for material on HIV/AIDS in southern Africa in a database.

---

[2]Paul Goldstein. 2001. *International Copyright: Principles, Law and Practice,* 72, Oxford University Press, Oxford.

[3]Stephen Ladas. 1938. *The International Protection of Literary and Artistic Property* 365, quoted in Paul Goldstein. 2001. *International Copyright: Principles, Law and Practice*, Oxford University Press, Oxford.

[4]See Chapter 6 of these *Proceedings*, "Overview of Legal Aspects in the European Union," by Thomas Dreier.

Professor David Vaver of Oxford University has written that:

> Intellectual property law as a whole seems ripe for wholesale reconsideration, both nationally and internationally. One might start with this fundamental premise that the system of rights it establishes enhances the goal of desirable innovation, creativity, and the widest possible distribution of ideas, information, products and technology in the most efficient and generally best way. The premise is, of course, empirically unprovable. We cannot, in fact, prove that the intellectual property system as it exists today produces the best innovation and creativity, but we all seem to think that we know that it is the best way. It is assumed that attaching a private property right to every activity with potential value in exchange, and thus creating a market in such rights, ultimately benefits not only the right holders but the communities in which they form a part. . . .[5]

This quote is a challenge to us all. We should challenge the presumptions of intellectual property, because, overall, intellectual property does not work for the benefit of countries of the South in the current conjecture.

[5]D. Vaver. 1997. *Intellectual Property Law: Copyright, Patents, Trade-marks* 270, Irwin Law, Concord, Ontario, Canada.

8

Economic Overview of Open Access and the Public Domain in Digital Scientific and Technical Information

Robin Cowan
University of Maastricht, The Netherlands

SCIENTIFIC RESEARCH AS A PRODUCTION PROCESS

Economists think about production as a process that has inputs and outputs. In science the output is fairly clear: new knowledge—at least that is our hope. There are three kinds of inputs: (1) human input—people, their brains, their time, their labor, and so on; (2) physical input—the buildings, the computers, the laboratories, the equipment material, and so on; and (3) the information input. For our purposes it is the third input that is the most interesting.

Several things are meant by "information input." There are data, which are factual information created by somebody in the past. There is also the primary scientific and technical literature and other related information products. And there is background information or know-how, which is learned in graduate school, by reading papers and books, by talking to people, and by listening to people at symposia. These are all various forms of information inputs. They get put together and combined with various physical and other human inputs to produce new knowledge.

Economists say that if the prices of inputs in some production process fall, you should get more output. In the case of science this means that easier access to knowledge should create more scientific output. Easier access implies easier knowledge creation.

ECONOMICS OF INTELLECTUAL PROPERTY RIGHTS

Where do intellectual property rights (IPRs) come into this picture? The economics of IPRs are based on the idea that economic agents, or any agents really, respond to incentives.

The problem with knowledge is that it is a public good, or at least a quasi public good, which means that everyone can use the same piece of knowledge without depleting it. One person using the periodic table of elements does not prevent anyone else from using it as well. Thus, the table of elements is a public good. That means the originator of the table of elements has a problem in appropriating the value from it. Intellectual property rights provide authors and inventors with some means to convert the public good into a private good. For a firm inventing a new product or process "getting the benefits" means making profits.

We therefore can generalize: IPRs create incentives to create new knowledge. They permit the inventor of a piece of knowledge or information to control the use of it in various ways, but in so doing they hamper the use of it.

Control of knowledge implies control of mechanisms that prevent the unfettered use of it. Other people have trouble using it relative to uncontrolled knowledge. IPRs thus also reduce the ability of other people to use this work.

The economics of IPRs is about striking a balance between incentives to invent and use of the knowledge for innovation and wealth creation, thus between inventors and users. There is a tension, and all of the economic literature on IPRs concerns the resolution of that tension or finding the optimal balance. This provides an explanation of why copyrights are used in scientific publications. People can use the information and resulting knowledge, but they cannot use the expression of it. There is an old-fashioned view that scientists are above all this, that they publish because they love creating knowledge. There is some truth in that; however do not think in terms of profits, but rather in terms of getting credit. A firm gets credit for its ideas by making profits. Scientists get credit for their work by other means. Plagiarism destroys a lot of the incentives for scientists to create because it allows another person to claim credit for the knowledge. This other person might become famous. The balance between creating incentives to publish and using the knowledge, which is the basis of the economics of IPRs, applies to science as well.

We might think that we have had a pretty good balance between these two opposing goals. Recently, however, two things have changed that have upset the balance. They are the information and communication technology (ICT) revolution and the changing financial situation of open science. I can only speak from my own familiarity with the English-language sources and knowledge of the developed world. The ICT revolution began to take off with the Internet, and accelerated with the World Wide Web. Financial changes began in the United Kingdom with Mrs. Thatcher and swept through the rest of the English-language world—Canada, the United States, Australia, and New Zealand—and have begun to affect Europe with a vengeance.

IMPACT OF INFORMATION COMMUNICATION TECHNOLOGIES ON INTELLECTUAL PROPERTY RIGHTS

What have the new ICTs done to change the balance? The new ICTs are different techniques for storing, searching, sorting, and rearranging information. These are all technologies that scientists need. We need to be able to store information. We need to be able to have access to it. We need to find it, rearrange it, and package it in different ways. The new ICTs have made all of this much cheaper and faster. That means a piece of information as a public good or input to science becomes much more valuable because of the way ICTs have changed our ability to access it and store it.

These changes in ICTs and the way information can be processed means that productivity in science should improve. We should be getting more results. On the other hand, such results are harder to control. If you are the originator or owner of a piece of information it is much harder for you to control that information if it is open, because it can be spread so much faster in so many different ways. This implies that the IPR regime is effectively being weakened.

While the laws have been strengthened, they have become much more difficult to enforce and this provokes additional calls to strengthen the regime. This is all part of a long trend toward easier and cheaper duplication. Consider a plausible but fictitious tale: When my father was a student, to find out how to integrate a particular function he went to the library, found the *Chemical Rubber Handbook*, and wrote down the integral of that function on a piece of paper. When he became a professor, to do the same operation he walked to the library, photocopied one page from the *Chemical Rubber Handbook*, and then walked back to his office. Obtaining the entire page was preferable to a single integral because he had the integrals of a couple more functions in case he made a slight mistake and needed a slightly different function. Some years later, when I was a student, the way it worked was that students went to the library and photocopied the entire *Chemical Rubber Handbook* so they would not have to go to the library again. Now that I am a professor, we do not go anywhere near the library. We go to Google, type in *Chemical Rubber Handbook*, and download the entire PDF file. This is a fable, of course, and fables have lessons. Understandably, the Chemical Rubber Company (or any firm seeing a similar reduction in its ability to control its product) would be upset by these changes in behavior. They no longer seem to be able to

control who uses the handbook. As soon as somebody puts a PDF file of it on the Web, they lose control, and in so doing they have lost their ability to make money from their product. It is perfectly natural that firms or individuals in this position would seek protection against people stealing their database.

Let me make a quick editorial comment. When the automobile was invented, the makers of horse-drawn carriages lost the ability to make money on their product. Did the public say, "That is pretty bad. What we should do is restrict the top speed of automobiles to 5 miles an hour" (actually this did happen but only for a short while) or "We should force automobile manufacturers to pay a royalty to the makers of horse-drawn carriages because their economic viability has been destroyed"? No, what we said was, "Too bad. Times have changed and we don't want your product anymore. There is a better one available. Find another way to make money." This is what the market said to the Encyclopaedia Britannica about 10 years ago when they were close to being bankrupted by the CD-ROM. Rather than calling for restrictions on CD-ROM encyclopedias they figured out a way to turn their product into something that would make money in the new world. It seems to me that the same sentiment should be told to the holders of databases instead of creating database directives.

FINANCIAL PRESSURES ON PUBLIC-SCIENCE INSTITUTIONS

Funding for open science or public science has been shrinking, at least in the North, for a couple of decades. As a consequence universities in particular, but research institutions in general, are looking at cost recovery; in other words, making things pay for themselves. Partly as justification for squeezing education budgets policy makers are saying things like "the problem today is that our scientists do not know what to do with their inventions. We know that there are billions of dollars of great inventions sitting on university lab benches, unexploited by industry. This is a colossal waste." This is the rationale for the Bayh-Dole Act in the United States and similar laws elsewhere. But there is a second rationale that has subsequently been adopted by university administrators as well: "We can solve the university funding problem just by getting the inventions out of the hands of the professors and into the market." Universities have succumbed to this temptation, and scientists are being forced to become entrepreneurs and to start making money on their inventions.

The idea is that there are many products being invented in universities and nobody, not even the university, is making money on them. There is pressure to change that, and universities are trying by creating technology transfer offices. As a result public research becomes a lot more like private research, especially with regard to intellectual property rights. There has been a huge increase in the number of university patents in the last decade. Universities are changing their views about IPRs and opting for much stronger protections on their inventions.

What does this mean for the future? I see two things that are both likely to happen, and neither is particularly nice. The first is that science will become harder or more expensive, which for an economist amounts to the same thing. The extent of protection for information goods increases, forcing scientists or their institutions to pay for information inputs, many of which were (monetarily) free in the past. If you want some data you have to buy them. If you want to use somebody else's technique, you will have to pay license fees. If you want some instrumentation, you will have to buy the instrument. This simply adds to the financial pressure on these institutions that then have to recoup yet more costs, thus creating a vicious circle. The incentives to make money out of everything we do become stronger and stronger, forcing private research institutions to make money by some other means.

The second is that the world of science divides into a group of haves and have-nots. In order to avoid the vicious circle just mentioned, consortia of universities will arise, the goal being to share their intellectual property among themselves. This is just what we see in the "patent pooling" of firms in the market. But not all of the universities in the world will get to join. Harvard, Stanford, MIT, and Chicago will probably form one consortium and Cambridge, Oxford, London School of Economics, and University College London will form another. Many other institutions will be left out. IPR sharing will exist, but naturally you want to share with someone who has lots to give you. While everyone might like to join the MIT consortium, MIT will not be that interested in having everyone join, as that would destroy the market for MIT's knowledge products.

CONCLUSION

Some external changes have upset the balance in the IPR regime in science. The new technologies make open science more valuable because of the way people are able to use increased amounts of information more effectively and quickly. This suggests that we should create more open information, which in turn suggests weakening the IPR regime. On the other hand, new technologies make it harder to control information products, which suggests that the IPR regime should be strengthened to give people back the control they are used to. At more or less the same time, financial issues create incentives to close information and to demand stronger protection for it. Given that these changes argue in opposite directions for what we should do with the IPR regime, it is possible that the right thing is to do nothing. But it is certain that the wrong thing to do is to react quickly to the lobby with the loudest voice.

9

Scientific Research, Information Flows, and the Impact of Database Protection on Developing Countries

Clemente Forero-Pineda[1]
Universidad de los Andes, Universidad del Rosario, Colombia

INTRODUCTION

This paper addresses the potential economic and scientific effects on developing countries of database protection and of the general trend toward restricting the public domain of scientific information. The analysis takes into consideration the specificities of scientific practices in developing countries and the institutional framework. Scientific knowledge and databases are produced differently in developing and developed countries. Some issues relating to differential access to scientific information are discussed. The economic and scientific impacts of granting further intellectual property protection to databases not covered by copyright agreements under the Trade-Related Aspects of Intellectual Property Rights (TRIPS) agreement are analyzed from this perspective.

The following questions guide this line of thought:

1. What is the role of information flows in science?
2. How is scientific knowledge produced in developing countries?
3. Is there any difference in the way scientists from developing countries use information compared with the use of information by their peers in developed countries?
4. How are databases produced in developing countries and for what kinds of markets?
5. Should developing-country scientists expect effects of *sui generis* protection of databases different from those envisaged for science in the developed countries?
6. What is the relevance of certain specific institutional arrangements, such as the public domain, "fair use," and differential or discriminatory pricing?
7. Are compensatory direct subsidies to developing countries viable?
8. What are the consequences of database protection and a narrower public domain of scientific information for international scientific collaboration?

Recent studies on the research process point out new roles of information flows and cooperation in the production of scientific knowledge. Although well recognized by writers on the social and institutional conditions

[1]The author acknowledges the dedicated research assistance of Abelardo Duarte. Carlos Iván Cañón and Daniel Escobar contributed to data processing. The author assumes sole responsibility for opinions and for any errors or shortcomings remaining.

of scientific communities, the role of the public domain in information is now being emphasized in the study of knowledge formation as a social, collective process.

An interesting approach to understanding the benefits of collaboration and fluid exchange of information is the evolutionary stance of Loasby (1999), who views a knowledge-rich society as "an ecology of specialists" that may grow provided it is "sufficiently coordinated to support increasing interdependencies" (p. 130). Nonaka and Konno (1998), in their attempt to explain the foundations of Nonaka's model of the conversion of tacit into explicit knowledge and vice versa, relate knowledge with the notion of *ba*, the Japanese word for a shared space for emerging relationships, and where knowledge conversions can unfold. Stacy (2001) has advanced the view that "knowledge is not a 'thing' or a system, but an ephemeral, active process of relating." Exchange and collaboration are central to this new epistemology, where flows of information are perhaps more valuable than stocks of information.

These three views converge into stressing the importance of a public domain in science and knowledge processes in general. They point precisely at the critical role of information flows for scientific activity in developing countries. Communication and collaboration are at the heart of these countries' aspiration to science, and access to scientific information is critical to their work.

INFORMATION AND THE PRODUCTION OF SCIENTIFIC KNOWLEDGE IN DEVELOPED AND DEVELOPING COUNTRIES

From an economic and social perspective the practice of science in developing countries is paradoxically akin to scientific research being performed in emergent areas of knowledge in developed countries. The replication of many simple experiments in developing countries usually requires creating a specialized and costly laboratory. Entering the circuits of world scientific activity or building networks of interested peers is accordingly difficult. Environmental, sociometric, and econometric research often demand initiating the collection of ad hoc statistical series. Scientists in these countries cannot easily reap the economic advantages of nonrivalry, in contrast with the belief of many economists who take for granted the nonrivalry of information, knowledge, and scientific networking. In this respect, science in developing countries is akin to frontier research carried out in developed countries, even when the purpose is obtaining a marginal result.

Based on detailed empirical studies by H. Collins (1985) and others, Michel Callon (1999) makes a distinction between two categories (or dynamic states) of science: emergent and consolidated science. In emergent science (science being done in emergent fields of knowledge) the diffusion of knowledge is not without cost. The transfer of knowledge (through replication for instance) demands heavy investments on the part of peers. Measurement instruments, seminars, and sometimes the building of new laboratories are necessary for knowledge to be appropriated. In contrast when scientific activity in one field is in a state of consolidated science (a concept akin to Kuhn's normal science), the traditional assumption of cost-free transfer of knowledge is acceptable. In the dynamic perspective proposed by Callon one scientific field may alternate between phases of emergent and consolidated science.

In certain aspects science in developing countries resembles emergent science. The replication of all experiments demands heavy investments. Transfers of knowledge from international peers require important local capital investments in scientific equipment, just as in the projects of their emergent science peers of developed countries. The percentage of equipment in project budgets is high. The process of networking and gaining the interest of their peers[2] is very demanding.

In other respects there are similarities between consolidated science in developed countries and science in general in developing countries. Expected results of projects are in general normal science propositions, developed within dominant paradigms or confronting marginal aspects of these paradigms. In both cases the reliance upon information resources is critical.

[2] "Building the universality of scientific propositions resembles more a public works enterprise than the miraculous conversion of spirits that are taken by evidence and the strength of logical arguments" (Callon, 1999, p. 35).

There are great differences as well between the normal practice of science in developing countries and both emergent and consolidated science in the developed world. Even the replication of simple experiments demands high investments because of the sparseness of scientific infrastructure, usually built in response to the demands of a few previous projects of each laboratory. The structure of expenditures in research projects shows that simple research projects in developing countries grant a share to library and digital resources, documentation, travel, and communication activities in a proportion that is considerably higher than that observed in emergent science projects of the developed world (CTS-Columbia Project, 2003). At the same time, access to financial resources is scarce.

Independent from the institution-building efforts by universities, public libraries, and special documentation projects in developing countries, it is common that a single research project will have to build information resources practically from scratch. For example, journal collections frequently are incomplete and subscriptions are restored only when individual projects demand it. This is the case when a research group or laboratory is built to nest a returning, recently graduated young researcher or when a team of experienced scientists joins to open a new avenue of interdisciplinary research. It is also true for new projects of well-established research teams. As a consequence, though reliance on information is equally critical for scientific activity in both developed and developing countries, a higher share of resources must be allocated in the latter to information resources to compensate for the sparseness of library and documentation resources.

The benefits to global science of incorporating researchers from developing countries have been analyzed elsewhere (Forero-Pineda and Jaramillo, 2002). Under a wide range of circumstances scientists from developed countries have shown interest in the research being done by researchers from developing countries. Both academic interest and altruism of scientists in the more developed world have played a role in meeting part of the information requirements of research projects in developing countries. In the basic sciences research projects in developing countries almost invariably refer to international links with laboratories and researchers in other countries. Scientific collaboration is the rule for leading research groups and fair use of copyrighted information is an important mechanism facilitating the access of researchers from developing countries to scientific information not directly available to them (CTS-Columbia Project, 2003). Some developing and transition countries have at times created networks of expatriate scientists that have played a key role in mobilizing scientific altruism (Forero-Pineda, 1997).

Besides the general argument showing how critical access to information resources is for scientific endeavors in developing countries, there are reasons to believe that scientists in developing countries may find comparative advantage in information-intensive research, provided access to information is maintained or enhanced. Consider the high cost of scientific equipment and reduced budgets for science in countries of lower income. Under these circumstances scientists in these countries would normally tend toward methodologies that rely specifically on open access to information. The types of research that depend more on the availability of and open access to information than on investments in scientific equipment are theoretical research in all disciplines; evidence-based research, both in medicine and in social practice; environmental, climatic, econometric, and genetic research based on internationally available data; and case-based or best-practices research in management or in technology applications. These types of research offer advantages to developing-country scientists. As the cost of buying and maintaining equipment is higher than that of maintaining well-informed scientists, and developed countries face less cost constraints, developing countries may have comparative advantages to develop these information-intensive types of research; even if their resources are scarce, the cost of supplying them with information resources is lower than buying and maintaining equipment.

OBSTACLES TO ACCESS

The scarcity of information resources and sometimes of the knowledge necessary to use them still affects the conditions of science in some developing countries. In others this situation has improved considerably as a result of three factors: considerable networking efforts by their scientists, altruism of their peers in the developed world, and the existence of rules (such as the exception of fair use in copyright) allowing and sustaining this cooperation. These rules allow information and other resources to be readily available throughout scientists' personal networks, and mitigate the handicap of developing-country scientists. They also allow a more important contribution of the

latter to the objectives of the common research programs of those networks. Considerable efforts have been displayed by some developing countries to strengthen these networks, sometimes supported and mobilized by metanetworks of expatriate scientists (Forero-Pineda, 1997). Nonetheless, legal and economic obstacles to information flows remain. The trend toward stronger protection of intellectual property rights over information that is deemed necessary for the production of knowledge, combined with the concentration of databases in developed countries, poses serious obstacles to scientific information flows for researchers in developing countries.

Concentration of Database Production

The production of databases is highly concentrated in North America and Western Europe; 94.1 percent of 12,111 databases traded worldwide and listed in the Gale Directory of Databases are produced in these areas (Williams, 2002, as cited by Braunstein, 2002). These figures do not cover all existing databases, only those that are commercially traded. In the case of databases from South America, for instance, the Gale Directory lists 21 databases, in contrast with 694 reported by Pereira et al. (1999) for Brazil in 1996. Many of the latter are governmental, legislative, and bibliographical databases and are not offered in international trade. As illustrated by this inventory for Brazil most databases produced in developing countries are of local interest. Many of them are not digitized, and most are provided at no charge. In the case of India a similar underestimation on the part of the Gale Directory seems to have occurred. It lists 413 databases for the whole Asian continent, including Japan, while Vandrevala (2002) mentions between 400 and 500 scientific databases in India alone. While the Gale Directory aggregates are perhaps an indicator of the concentration of internationally traded databases and reflect international payments flowing from South to North, they appear to ignore a large majority of databases in both developed and developing countries.

Network Access

There are network limitations for developing countries to access the production of databases that come from the absence of network economies in these countries. Internet networks in these countries are thin and small. Their cost has to be shared among fewer users, making them more expensive.

A Lorenz curve may be built, based on Kirkman and Sachs (2001) data on network readiness and United Nations statistics of world population, showing the concentration of Internet access (see Figure 9.1). Eighty-five percent of Internet users are in countries with only 20 percent of the world population. Figure 9.2 shows the size of the problem and the disparity of conditions in terms of income and Internet access. While income was not well distributed in the world in 2001, it was better distributed than Internet access.

The average size of telephone networks in the 75 developing countries studied by Kirkman and Sachs (2001) is 8 million telephone lines. In Organisation for Economic Co-operation and Development (OECD) countries it is 20 million lines. The average size of the Internet network is 1.5 million telephone lines in developing countries compared with 12 million telephone lines in OECD countries.

There are important economic consequences of this statistical analysis. Developing countries have less opportunity to take advantage of network economies and therefore less potential for database development. Though database markets appear to be global, database development is linked to data production, and both are concentrated.

THE LEGAL PROTECTION OF DATABASES AND SCIENTIFIC ACTIVITIES IN DEVELOPING COUNTRIES

Original and Nonoriginal Databases

Among the varied information resources useful in scientific research that may be included in the wider definitions of databases, digital databases are extremely important for international scientific collaboration, since the Internet enables the immediate transfer of pieces of information. The debate on providing additional statutory protection for databases has centered on *sui generis* protection of nonoriginal databases, since original databases

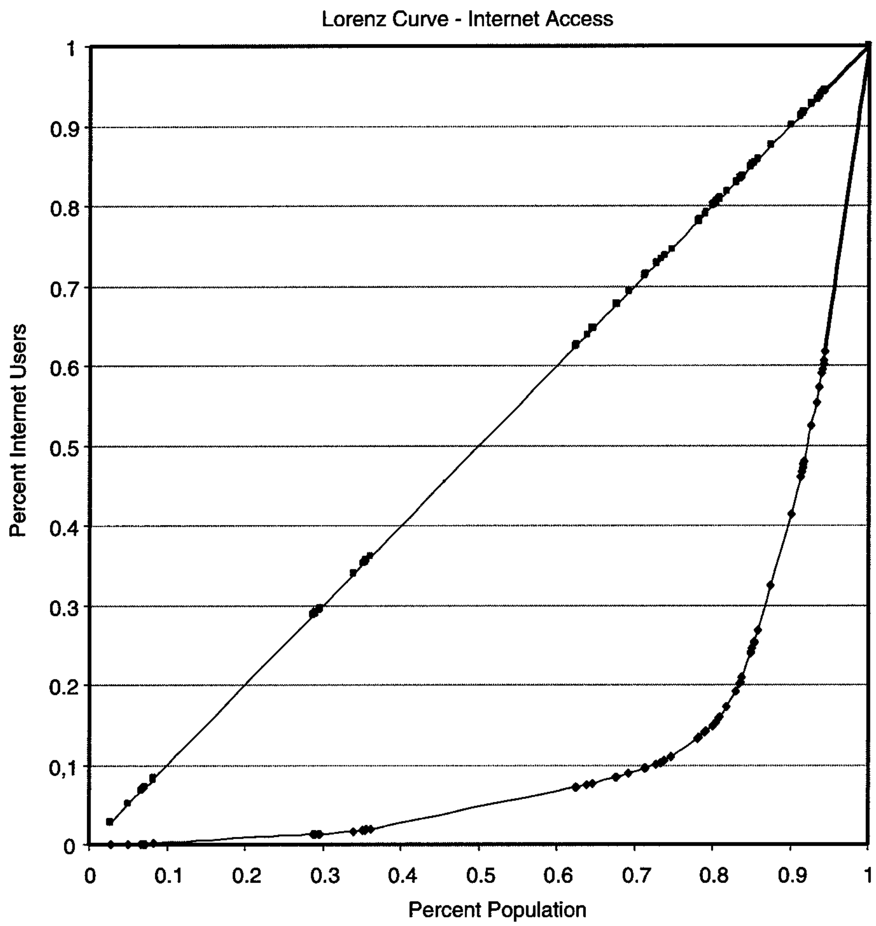

FIGURE 9.1 Worldwide Internet access.

are already protected through copyright, with special exemptions including fair use for academic purposes. That focus, however, should not preclude consideration of the importance of databases in scientific activities of developing countries. The arguments presented previously apply to both.

The border between these two categories is not always sharp, as many authors recognize but often dismiss in their theoretical thinking. Database developers frequently integrate their own information with information available in the public domain. If additional protection is granted to databases already protected by copyright, the exceptions for fair use and the de facto perpetual protection that is granted to databases in certain regimes, equally affect the supply of information for research activities in developing countries.

An issue of special concern in developing countries is that of legal protection of nonoriginal databases. While the TRIPS agreement allows protection of original databases, the protection of nonoriginal databases is granted in the European Union, some northern European countries, Mexico, and South Korea. A significant concern is that even information that is in the public domain simply could be reorganized or versioned and included in proprietary databases.

Preliminary analyses made for Latin American countries show that local production of nonoriginal databases does not seem to be affected by the absence of legal protection (López, 2002), and that most nonoriginal databases are produced elsewhere. This could explain the reluctance of most of these countries to support additional legal protection of databases in the WIPO discussions of 1996.

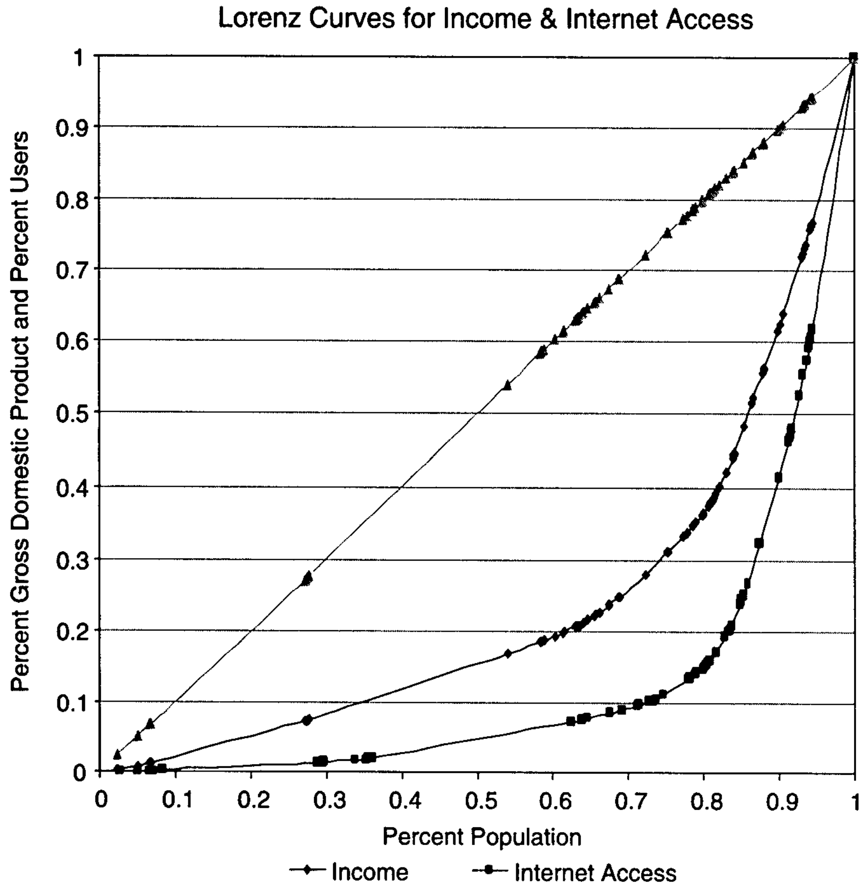

FIGURE 9.2 Income and Internet access.

Information as a Public Good and Information in the Public Domain

The border between private and public-domain information is not well defined. From an economic point of view a distinction should be made between information as a public good and information being in the public domain. Information may be viewed as a public good when nonrivalry applies to its use and it is difficult to exclude others from its use. This is a category that covers most information. As a corollary a piece or set of information may be viewed as a global public good when nonrivalry and nonexcludability apply to it on a global scale.

In contrast, the public domain may be analyzed in an institutional setting or arrangement. Information is not in the public domain by its public good nature or even by its governmental origin but as the result of a network of formal and informal social agreements, explicit or implicit but entrenched in common law and in the culture of a society. There is a tradition ascribing most information obtained in government-funded science to the public domain, with the important exception of classified research for purposes of national security or economic competitive strategy. Nonetheless, the public domain for information generally is wider than that. It includes information that has lapsed protection, information produced by intergovernmental agencies, or information "contractually designated as unprotected," just as a legal definition may state. From an economic perspective an undefined area still remains. In the case of information produced by a private firm, data might be gathered by an explicit, intentional, and marginally costly effort on the part of the firm; alternatively, data collection may be "incidental"

to a public activity. For instance, it could be related to a government-granted concession for providing a public service or in general be a virtually costless by-product of other public activities. Should exclusive exploitation rights be granted to the firm over this information? Or, extending the principle and tradition of ascribing government-agency produced information, should it be considered as part of the public domain? The implications of defining these borders as well as those of the definition of databases protected under *sui generis* legislation are important (Maurer et al., 2001).

Pricing Methods, Transaction Costs, and Access

The market for industrialized countries' databases in developing countries has very specific characteristics. As in all network products there is an incentive to use introductory pricing or discriminatory pricing.

Differential pricing is a common practice in the database market. Databases may have a posted price of U.S. $50,000 and sell for U.S. $20,000 to a consortium of universities in one developing country and for U.S. $6,000 in a smaller country as an introductory price. As predicted by both the theory of discriminating monopolies and the theory of networks, the pricing rule applied by intermediaries appears to be directly related to the income of countries and to the expected volume of use of the database.

Database trading in developing countries is also subject to high transaction costs. Besides the price charged by database owners there is a dense network of intermediaries operating in developing countries. In the Andean region of South America there are at least five permanently based intermediaries, who operate through representatives. One more intermediary acts in this region through a subsidiary in Brazil.

The strategy used by universities in developing countries to deal with the problem of the cost of these databases has been relatively successful. These universities have created consortia to negotiate access to databases. The perception of academic consortia of database users is that negotiations with intermediaries are considerably harder than dealing directly with database owners.

On the supply side it should be observed that incentives for the production of new database firms increased sharply for one to two years after the European database directive according to Maurer, Hugenholtz, and Onsrud (2001). However, in developing countries, as López (2002) states, ". . . we have not observed that infant DB industry in the region (apparently concentrated in the most advanced countries) is being affected by the absence of *sui generis* legislation. Commercial losses seem instead to derive from an inadequate enforcement."

There is concern that if protection were extended and public agency databases were transferred to private dealers, price increases in developing countries could be substantial. According to López existing protection schemes in these countries are being effectively used, as observed in court cases, and the problem for the development of this industry is more one of enforcement and costs of legal processes than one of insufficient protection.

Closely linked to the issue of prices is that of exclusive use. If there is complete freedom for pricing, the level of the price might be so high as to preserve an exclusive or almost exclusive use, discriminating through the buyers' ability to pay. Access is clearly affected by these monopoly practices, whether illegal or compliant with the law. Misuse of exclusive rights could be confronted by antitrust litigation, but such legislation is uneven in developing countries and transaction costs of suing in developing countries are very high. The risk of being condemned through litigation is therefore low and it would hardly dissuade monopoly practices in developing countries.[3]

The economic and scientific consequences of the legal protection of databases on the scientific and technological activities of developing countries should be analyzed in this perspective. Additional legal protection might stimulate new information businesses, in both developed and developing countries. Nonetheless, a simple market analysis would predict that greater legal protection will make access to protected databases more expensive on the average. The incentives for the production of open-access scientific information will accordingly diminish. If, as

[3]López (2002) has mentioned additional concerns related to *sui generis* protection: the issue of pass-through rights, the elimination of fair use, and the de facto perpetual protection of databases in the model of the European Union Database Directive. The negative consequences for international scientific collaboration are easily deduced.

shown by Kirkman et al. (2002), access to scientific information is today more expensive for residents of developing countries, these shifts tend to further reduce the possibility of access by scientists in university and government research institutions in developing countries, especially those where university and public library budgets have shown a sharp reduction in the past decade.

CONCLUSION

The arguments presented here are based on a stylized analytical description of how science is produced in developing countries. Results are similar to those of authors who have analyzed the impact of database protection in an institutional context. They differ, however, from those of other authors who present the classical arguments in favor of a strong protection without taking into account the specific ways of producing science in developing countries.

If one adopts the view that institutions and the production of science matter, then the fair use exception of copyright over original material as well as nonoriginal databases; the disclosure provision of patent law; and effective limits to the time span of property rights appear to be fundamental to the advancement of science in developing countries.

If the observed trend toward stronger legal protection of databases were to advance, the end result might not be the end of scientific collaboration, as some have claimed. A new topology of scientific research networks would probably emerge, however. Rather than a global commons for scientific information, a collection of closed and relatively isolated networks would dominate scientific activities worldwide. In this scenario the role of researchers from developing countries in global science would diminish even further in relative terms, as a consequence of the narrower availability of scientific information.

REFERENCES

Braunstein, Y. 2002. "Economic Impacts of Database Production in Developing Countries," WIPO, November.

Callon, M. 1999. "Réseau et Coordination," *Economica*, Paris.

Collins, H. 1985. *Changing Order, Replication and Induction in Scientific Practice,* Sage Publications, Thousand Oaks, CA.

CTS-Colombia Project. 2003. "The impact of science and technology on Colombian Society," Universidad de los Andes, Universidad del Rosario, OCYT, Colciencias, Progress Report, September 20, Bogotá.

Forero-Pineda, C. 1997. "Convergence of research processes, big and small scientific communities," *Proceedings of the Global Science System in Transition*, Austria, May.

Forero-Pineda, C., and H. Jaramillo-S. 2002. "The access of researchers from developing countries to international science and technology" (coauthor), *International Social Science Journal*, Special Issue on the Economics of Knowledge, No. 171, February.

Kirkman, G. S., P. Cornelius, J. Sachs, and K. Schwab, eds. 2002. *The Global Information Technology Report 2001-2002: Readiness for the Networked World*, Oxford University Press, New York, April.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions,* University of Chicago Press, Chicago.

Loasby, Brian. 1999. *Knowledge, Institutions and Evolution in Economics,* Routledge, London.

López, A. 2002. "El Impacto de la Protección de las Bases de Datos No Originales sobre los Países de América Latina y el Caribe," World Intellectual Property Organization, Geneva, November 4-8.

Maurer, S., P. B. Hugenholtz, and H. J. Onsrud. 2001. "Europe's Database Experiment," *Science*, 294, October.

Nonaka, I., and M. Konno. 1998. "The Concept of Ba: Building a Foundation for Knowledge Creation," *California Management Review*, 3(40):40-54.

Pereira, M. N. F., C. J. S. Ribeiro, L. Tractenberg, and P. Loureiro Medeiros. 1999. "Bases de Dados na Economia do Conhecimento: A Questão da Qualidade," *Revista Ciência da Informação*, Volume 28, No. 2.

Stacy, R. 2001. *Complex Responsive Processes in Organizations: Learning and Knowledge Creation (Complexity and Emergence in Organizations)*, Routledge, London.

Vandrevala, P. 2002. "A Study on the Impact of Protection of Unoriginal Databases on Developing Countries: Indian Experience." *National Association of Software and Service Companies (NASSCOM),* WIPO, Geneva, November.

Williams, M. E. 2002. *The State of Databases Today: 2002*, Gale Directory of Databases, Gale Research Inc., Detroit.

10

Information Technology and Data in the Context of Developing Countries

Chrisanthi Avgerou
London School of Economics, United Kingdom

Although my affiliation is the London School of Economics, I am not an economist. People assume that the London School of Economics is populated by economists, but that is not the case. I am in information systems, where we study the efforts required to make use of new information. That makes me appreciate the significance of information and communication technologies, as well as the significance of information and data.

It is not enough for organizations to have access to technology or to be able to acquire technology. It is not enough for organizations to have information and data assets, whether internally or externally generated. They also should have the capacity to use all of the technology and information resources to their own advantage. Nevertheless, in our efforts to emphasize the significance of access to data and information in the current discourse of the global information economy and the digital divide, we run some risks. We may neglect some important issues that do not appear significant at the beginning.

I will briefly discuss the risk of reifying data, information, and knowledge. Despite the definitional differences of these three concepts, it is problematic to consider data, information, and knowledge as objects possessing some value that is independent of context and of other aspects within which they are embedded. I will build on that concept and argue for the need for analytical capabilities over and above the need to have access to data and information. In relation to that I will argue for the significance of social sciences, which in my opinion, are weak in many parts of the world.

THE RISK OF REIFYING DATA

The risk of reifying data is the problem of assigning value, assuming that there is value in data and information, without consideration of institutionalized practices, both in terms of where the data come from and also where these data or information are going to be put to use. There is now a whole field, the sociology of science and technology, that has created awareness about the process of scientific and technological development. This process by no means discovers the truth, but rather ends up with packages of knowledge, which tend to be black boxed, and with alternative contested knowledge, which tends to be forgotten with time.

When we talk about public access to scientific data and information, we tend to forget that these resources contain codified information that has acquired legitimacy in specific historical and socioeconomic contexts; at the time of their production there were different research projects or alternative areas of science that for whatever

reason were marginalized. The data from these alternatives may still be relevant and in some cases more relevant in the particular context where the scientific knowledge will be used.

The institutionalized context within which scientific knowledge is produced can be very important. It is also very important to be able to relate whatever data, science, or knowledge you have with the prospective problem area where the knowledge will be used.

There is a considerable effort to emphasize the significance of science in developing countries. It is important to be able to unpack the black boxes of scientific data and be able to interpret them, understand their contextual nature, and make choices that are contextually appropriate. There is, however, something else that we tend to forget, which is the significance of scientists from developing countries being able to participate in the scientific debate worldwide.

I would like to share with you a very unpleasant experience I had as the editor of a special issue of a prestigious scientific journal. The special issue was supposed to cover the area of information systems in developing countries. I was an organizer of a conference in 2000 that was convened under the umbrella of the International Federation for Information Processing in the Developing Countries Group. For a number of years the federation had established a tradition of being quite open and inclusive in its participants. The conference organizing group considered it very important to have participants from all over the world to be able to exchange information.

The conference was successful and it was decided that a special issue with the best papers presented would be produced. The papers chosen were from authors from both industrialized and developing countries and went through peer-review processes. Unfortunately, two colleagues from developing countries were not able to include their papers in the special issue. Why? Scientific processes are very much institutionalized; there are rituals of writing, referencing, and arguing that are very much a part of our academia. Developing countries very often do not have the capacity to take part and prove their point as competently in these rituals as peer reviewers would like to see. In the end the special issue had only authors from industrialized countries. One lesson to be learned from this is that we should develop the institutions of science in a broad-minded way and perhaps be prepared to change our rituals in order to accommodate more voices.

THE NEED FOR ANALYTICAL CAPABILITIES

There is also the problem of data acquiring objectivity and universal truth status. There are philosophers and sociologists of science who have argued about the quality of scientific knowledge—that it might be reassessed, that alternative knowledge within science might prevail. Yet, the current perception of scientific knowledge is that it tends to take for granted such knowledge as universal truth. Not only is scientific knowledge not universal truth, but it hides power. By that I do not mean the rather clichéd view that knowledge is power; I mean the opposite. What we take as scientific knowledge, as prevailing truth, is very much an output of the dynamics of power, as I think my example illustrated. My colleagues from developing countries, not having mastered the conventions of writing their ideas in a way that is acceptable within the current norms of the scientific community, lost their voice. Their points simply were not included in a special issue on the very topic of information technology and development.

I would like to argue for the need to develop analytical capacities to be able to interpret data and make critical judgments about the validity of data in specific contexts. This is not a trivial issue. You should be able to unpack what is already black-boxed and have the ability to argue for your choices, often against the prevailing legitimate practice.

There is also the need to juxtapose alternative context-specific knowledge. For some time there has been a debate within development studies about the relative status of scientific knowledge and indigenous knowledge, an unfortunate dichotomy. It is very important to accommodate alternative knowledge within one's own epistemology and underlying values.

Again, this requires an ability to examine the reified notion of data and information and develop critical judgment about what is relevant, beneficial, and feasible. The development of policies that give a voice to the public and consider their own judgments of issues of relevance to their lives is weak in developing countries.

Another weakness of developing countries is social science. Some areas of social science are very much the Cinderella in academia because they are not tangible and do not require labs or technological infrastructure. It is even slightly dangerous to the local elites because it raises some difficult questions. It encourages inquiries into values, thus quickly entering politics.

However, the division between science and value judgments or political choices is blurred more and more, making it important to develop scientific abilities in social issues. Social sciences in many respects are different from the natural sciences. They are much more discursive. They require a totally different way of thinking and analyzing that complements that of the natural sciences.

I would now like to give an example that illustrates the blurring between data and value judgment, or perhaps judgments and assumptions that have been codified into data (see Table 10.1).

Such data and assumptions need examination. These data make associations between the availability of technology or the capacity to use technology and the economic development index. The data were produced by a group of scientists, economists, and information technology experts at Harvard University's Center of International Development. This illustration shows what we assume and accept as obvious—that industrialized countries, those countries that have more technology and more capabilities to use technology, are the more economically advanced countries. There is also the national readiness index rank. The countries at the top of this index are the richest countries while the poorest have very little capability with technologies. While we are all familiar with that correlation, the difficulty is interpreting what this means. The dominant interpretation at the moment is that information technology and the capabilities for using it, in other words, the readiness for information society in terms of technology, are tools for development.

It could be the opposite as well. It could be that those countries that are already advanced economically and technologically are in a better position to use their technology to further their economic growth. The interesting tricky question here is, if we take the view that technology and information are related to the capacity to analyze information, in what way is information technology a tool for development? For example, it is acknowledged that for the top tier of rich countries, information technology in terms of innovation becomes a competitive weapon and a factor for further economic growth.

Of course we know that information technology can be significant for development in many other ways. Nevertheless, a lot of examination, analysis, and critical judgment are required to decipher these data and make sensible use of them.

TABLE 10.1 Network Readiness Index

Country	NRI	NRI rank	Country	NRI	NRI rank
United States	6.05	1	China	3.10	64
Iceland	6.03	2	Romania	3.10	65
Finland	5.91	3	Ukraine	3.05	66
Sweden	5.76	4	Bolivia	3.04	67
Norway	5.68	5	Guatemala	3.00	68
Netherlands	5.68	6	Nicaragua	2.83	69
Denmark	5.56	7	Zimbabwe	2.78	70
Singapore	5.47	8	Ecuador	2.65	71
Austria	5.32	9	Honduras	2.64	72
United Kingdom	5.31	10	Bangladesh	2.53	73
New Zealand	5.23	11	Vietnam	2.42	74
Canada	5.23	12	Nigeria	2.10	75

SOURCE: Adapted from G. S. Kirkman et al., 2002, *The Global Information Technology Report 2001-2002: Readiness for the Networked World*, Oxford University Press, New York.

SESSION 2: DATA AND INFORMATION IN THE PUBLIC HEALTH SECTOR

11

Introductory Remarks by Session Chair

Dialo Diop
Université Cheihk Anta Diop, Senegal

In developing countries it seems appropriate to use the public health sector as a yardstick for assessing both the opportunities and challenges of open-access and public-domain scientific information. In terms of public health policy making, in a context of limited resources, the ongoing information and communication technology revolution offers many opportunities, such as the easy management of epidemiological data and biomedical statistics and distance learning and training of professionals. These opportunities have been reviewed extensively.

The challenges faced by developing countries, mainly in Africa, should be stressed. First, third-world researchers have encountered many problems producing and validating original data before processing them to generate information, which in turn must be organized and analyzed, to yield knowledge. Second, these researchers also must cope with technical constraints, such as poor energy or telephone infrastructure, along with financial restrictions of computer equipment or software availability. Once these obstacles are overcome, they still face market constraints for publishing and disseminating their own products. The public health sector also faces the current fierce debate about the patenting of living organisms' sequence data (e.g., human genomics), in which the major objective remains private profit.

That is to say, public health professionals in developing countries suffer diverse handicaps for both accessing and producing useful biomedical knowledge, whether electronic or printed. This lack of access poses an enormous threat in a knowledge-based global society, especially in this vital domain where curative and preventive medicine tends to become more and more predictive. In terms of patrimony, this trend accounts for a great loss to the universality of medical knowledge creation. In that respect, the relevance of traditional therapy and indigenous knowledge is conspicuous.

During Session 2 the presentations will address health-related success stories of attempts to narrow the North-South digital divide in Africa, Asia, and Latin America. The main lesson to be learned from these examples is the tremendous local impact of initiatives requiring only minimal technical and financial input.

In the information age it is commonplace to say that knowledge means power. Since globalization is unequal and asymmetric, there are legitimate concerns about the concentration of the means of domination (both material and immaterial) at the center of the "global village" and the complete marginalization of its periphery. This perspective is unacceptable as far as public health is concerned, not only because among all basic human rights good health is a prerequisite to the effective implementation of any other right (e.g., freedom, education, information) but also because health and knowledge as public goods are incremental processes—assets that increase rather than decrease upon sharing, individually and collectively.

In other words, since exemptions to the intellectual property rights regime are already conceded for education and research purposes, they should be extended to the public health and environment domains as well. Unrestricted availability to public information should be allowed because upstream information and scientific knowledge is mainly publicly funded.

If the global North-South divide is to be overcome in any field of human society, the unrestricted and free access to public information and the primacy of sustainable human development benefit over short-term individual profits are required. This appears to be the only way to convert threats into opportunities and switch from a zero-sum game to one that is positive. This requires developing countries to join together to protect and enlarge the public domain and promote open access. The crucial obstacle in this respect remains the lack of political will of the global leadership.

12

The Ptolemy Project: Delivering Electronic Health Information in East Africa

Massey Beveridge
University of Toronto, Canada

There is considerable discussion about the importance of access to electronic health information in poor countries, but to date there have been few projects that deliver the information directly to the doctors who need it and that evaluate how useful they find it. The Ptolemy Project[1] provides free access to the University of Toronto Library's extensive full-text electronic resources to 100 surgeons in developing countries, mostly in East Africa, and uses periodic surveys to assess their satisfaction with the information they receive. None of us knows how the dissemination of electronic health information will take place in the developing world in ten years time, but a safe guess is that there will be a number of different models, with each adapted to a particular niche, and until we experiment with different models we will not learn what works best where. BioMed Central, the International Network for the Availability of Scientific Publications (INASP), Scielo, the Health InterNetwork Access to Research Initiative (HINARI), Ptolemy, and Bioline (to name only a few) all represent different but complementary approaches to delivering electronic health information from local, regional, and international sources to the clinicians, teachers, researchers, and policy makers who need it. Six features distinguish Ptolemy.

1. It is university-library based, not publisher based.
2. It connects the medical literature directly to the end users, when and where they need it.
3. It makes scientific journals from the South available electronically, through its linkage with Bioline.
4. It has an ongoing, built-in evaluation mechanism.
5. It engages African surgeons in an online research community and encourages North-South research partnerships.
6. It could easily and economically be replicated by university libraries anywhere.

BUILDING RESEARCH COMMUNITIES

The health problems of Africa will be solved in Africa by Africans, for it is they who know the right questions to ask to get practical answers and find solutions that work. Yet how are they to do meaningful research if they

[1]Information on the Ptolemy Project is linked to the University of Toronto's Office of International Surgery Web site at http://www.utoronto.ca/ois/.

cannot read what has previously been written on their topic? Ptolemy directly addresses the 10/90 gap in health research by providing access to the literature for a group of surgeons in East Africa that has considerable leverage despite its small size. The membership of the Association of Surgeons of East Africa comprises 400 surgeons who are responsible for 200 million people in the eight East African countries; of those 400, 100 are Ptolemy participants.

Why is surgery so important in poor countries? The World Health Organization devised the Disability Adjusted Life Year to describe the global burden of disease. In this index, injury accounts for 12 percent of the global burden of disease, more than HIV, or diarrhea, malaria, and TB combined.[2] There is a global injury pandemic and in 2001 injury killed just over 5 million people around the world yet world attention remains focused mostly on other health issues. Those who are killed or disabled by injuries would usually benefit from surgical management, and poor outcomes can be related to the lack of surgical care. Ptolemy is a very direct and simple way to improve clinical practice, teaching, and research capacity in Africa and hence the ability to reduce the burden of injury there.

When considering how to build research capacity in Africa, one must remember that doctors are not paid to do research. Doctors are paid to take care of patients, operate, run clinics, and do clinical work. Why do doctors conduct research? In Western countries there exists a system of intangible rewards whereby research work is rewarded with the respect of colleagues and status in their communities. While this system of intangible rewards exists in industrialized countries, it is not present in much of Africa. In order to build research capacity we must build a community of medical curiosity that engages doctors mentally and draws them into the endeavor of doing research. How do we engage African doctors in such a community of medical curiosity? To change behavior we have to get the right information to the right persons at the right time and create the system of rewards that will persuade them to stay up at night working on a research project rather than spending that time in other activities. We have to see that their work gets published and that what has been published is indexed and widely available: only then can we expect to see competition-based research funding begin to flow to African researchers.

ORIGINS OF THE PTOLEMY PROJECT

The origins of the Ptolemy Project are at the intersection of a surgeon's interests and a librarian's. The surgeon had returned from working abroad and recognized the boon that access to the medical literature would be for colleagues who had no access to it and the librarian saw the Internet as perfectly suited to serve as the medium for this transfer of knowledge. There is a series of arguments about the inequity of access to health care information and the necessity for building equality. Sharing access with African doctors is a good thing to be doing. It is beneficial to the university; it enhances its profile internationally and has spawned an interesting research topic about how people actually use the information. Thus was born the Ptolemy Project, named after Ptolemy Sotor, the general assigned by Alexander the Great to found the city of Alexandria, who built the great library there in the third century B.C.

In establishing Ptolemy we also had to consider publishers' contracts. After examining the wording of the agreements between the big journal publishers and the university library, we found that the university community was defined as "students, staff, faculty, and affiliated researchers." We therefore decided to establish a mechanism for making researchers in Africa "affiliates" of the Office of International Surgery researchers in order to connect them with our university and facilitate library access. The library has dedicated 100 proxy server accounts for this. The University of Toronto Library has 50,000 existing library accounts, so the number of accounts dedicated to Ptolemy is not large (0.2 percent). Library accounts provide online, full-text access to a total of over 20,000 journals (2,500 medical journals) and several thousand textbooks.

In order not to provide access to those who might be able to afford journal subscriptions independently, we restricted participation to those individuals practicing in countries rated lower than 65 on the Human Development

[2]World Health Organization. 2002. "Annex Table 3: Burden of disease in DALYs by cause, sex and mortality stratum in WHO regions, estimates for 2001," in *The World Health Report 2002: Reducing Risks to Health, Promoting Healthy Life*, World Health Organization Press, France, October 30.

Index. Participants must be sponsored by University of Toronto faculty. Participants also have to consent to monitoring of their proxy server activity; the project tracks which sites they go to while they are logged on to the site. They have to respect copyright laws the same way any other user of the library does. And because we are surgeons and believe that surgical care for people in poor countries is essential and has been sadly neglected, we give preference to our surgical colleagues.

THE PTOLEMY PROJECT

The objective of the project is, first and foremost, to provide a group of surgeons in East Africa with access to medical literature and to evaluate the impact this access has on clinical work, teaching, and research. We wanted to strengthen the existing medical community in East Africa and improve the methodology for analyzing how they used the resource. We hope to demonstrate eventually that improved access to information helps build research communities and in the end improves patient outcomes. The health information that is most useful in Africa may not come from the North, where journals have basic science papers and cover topics that may be irrelevant in Africa. What may be of more use for African doctors is to find out what research is being conducted in other developing countries, such as India or Brazil. Bioline International is a collaborative initiative of the University of Toronto Libraries, Canada (management office), the Reference Center on Environmental Information, Brazil (host computer and software development), and Bioline/UK (liaison). It started as a cooperation between Brazil and the United Kingdom to make the electronic publication of journals from the developing world feasible and is now housed at the University of Toronto Library. Bioline is available free through Ptolemy. We are in the process of publishing the *East and Central African Journal of Surgery* to add to the Bioline list. The Ptolemy Project not only provides access to journals in the University of Toronto Library to the developing world, but it also works to make electronic access to journals from developing countries possible. There is hope for an information exchange that does not flow just from North to South but rather flows South to North and perhaps most importantly of all from South to South.

Lives depend on surgeons knowing what they should know right now. The usefulness of the information is time dependent and surgeons are busy people. During the day they operate in the hospital, and most do their research, reading, and writing in the evening, so convenience of access is very important. The Ptolemy Project provides Internet access at home to participating doctors in Africa, most of whom report their home Internet connection is much faster than the Internet access at their institution.

A participant survey was conducted in April 2002 at which time Ptolemy had been operating for four months and was receiving almost 9,000 hits on the site every two weeks. Our survey asked whether access to Ptolemy information had changed their practice of surgery, and two-thirds of the participants replied affirmatively. When asked whether Ptolemy had helped their research projects, 75 percent said "yes."

By August 2002, 118 participants had joined and 21 had left mostly due to Internet connection problems. A problem we encountered in East Africa was a difficulty in integrating the local-area-network Internet connection with the university library local area network. This problem has recently been completely resolved following installation of new software at the University of Toronto Library, so access is by password rather than by proxy server. Of the 97 people involved in the study, 78 were in Africa (69 were from countries of the Association of Surgeons of East Africa and 58 were surgeons). It is interesting that 51 of these participants taught surgery and 42 did research. The study had a total response rate of 79 percent.

It is interesting that two-thirds of the participants spent more than one hour a week browsing Ptolemy. That is probably more time than most surgeons in Canada spend on reading. The median cost was about US$42 a month, which is very expensive by African standards. The fact that people are actually willing to pay this rate indicates how much they value the resource.

Only 31 percent logged on exclusively from their work environment. Typically in African hospitals and libraries there are often only 10 or 20 computers connected to the same land line. Internet access can be painfully slow through these institutions, whereas at home the surgeons generally have fairly high-speed access. Though it may be unreliable and expensive, it is available, and they are prepared to pay for it.

We also asked whether Ptolemy was relevant to their practice in three areas—clinical research, clinical teaching, and research. Eighty percent of the respondents thought that it was very relevant or relevant. We also asked about the impact of Ptolemy on their practice; once again a very high percentage found it either strongly enhanced or enhanced. Perhaps the effect Ptolemy is having in Africa is best expressed by the participants themselves (see Box 12.1).

BOX 12.1
Sample Comments from Ptolemy Participants

"Up until I joined the Ptolemy Project I was only using abstracts for my work. This made life difficult and my publication record to date is not good. The last good paper I published was in 2000 when I finished my Ph.D. I certainly regretted going back home to Africa as I thought my academic career was over. I now know that I will be up to date and I will certainly come up with innovative research proposals." Dr. J. C. (Malawi)

"I am writing my dissertation for Masters in Public Health and Ptolemy is assisting me at just the correct time." Dr M. S. (Tanzania)

"I am very much interested in medical education, especially clinical education. The Ptolemy Project helped me to find relevant information about the subject. It should be noted that, due to financial difficulties, we do not subscribe to any medical education journals." Dr. P. G. (Mozambique)

"I do my hernia operations differently just for starters." Dr. J. B. (India)

"Our library is poor in Lusaka, Zambia. No journals or books. Ptolemy opened for me new world of knowledge. Like a child in a toy shop. It is difficult to stop once you start browsing. Unfortunately Internet connection is often very poor and downloading or opening page is difficult. That is the time when I stop. Preparing lecture in Wound Healing I performed almost entirely from Ptolemy." Dr. G. J. (Zambia)

"I have been able to write my proposal on a burns unit using the information I obtained from Ptolemy." Dr. P. O. (Kenya)

"My research on areas of developmental biology has received tremendous boost particularly that one is able to get full articles as opposed to other sources such as PubMed where only abstracts may be available." Dr. R. M. (Kenya)

"I did not have access to most journal articles full text, and if I needed a paper badly, I would have to ask a colleague from Europe or North America to search, print and fax me the article. It would take forever and I could not use this method too often. Now if I want a paper I download it off my computer. The only hitch is a slow and unreliable internet link . . . and I don't have to go to the library where services are usually slow and crowded." Dr. O. K. (Uganda)

"The internet access in Gondar College of Medical Sciences, where I work is not good. Since the line is very slow, often I had difficulty of logging into your server. But during the limited time that I was able to access the server of the U of T, I was able to retrieve relevant materials which enabled me to draft a paper, which hopefully will be published in the *Ethiopian Medical Journal.* I am grateful to the Ptolemy Project for this." Dr. S. B. (Ethiopia)

". . . no digital divide as long as Ptolemy is there. I have been able to access all I need from the library." Dr. S. K. S. (Tanzania)

The cost of operating the Ptolemy Project is low because there is no incremental cost to the library of opening such a small number of library accounts. The only real cost is the salary of a part-time research assistant and the associated office expenses. Researchers from the Office of International Surgery in Toronto periodically travel to Africa and provide support for participants as a courtesy. The Ptolemy annual budget presently runs about U.S. $10,000 per annum.

Ptolemy delivers the information to the people who need it, when and where they can use it. Ptolemy draws surgeons into a virtual research community and builds communication links where there were none before. Ptolemy has a very low incremental cost, most of which is related to the research component of the project.

DISCUSSION

As previously mentioned we must experiment with different electronic health information models to learn what works best where. Initiatives such as BioMed Central, INASP, Scielo, HINARI, Ptolemy, and Bioline all represent different but complementary approaches to delivering electronic health information from local, regional, and international sources to the clinicians, teachers, researchers, and policy makers who need it.

HINARI is a unique and powerful coalition of high-level organizations (publishers, *British Medical Journal* [BMJ] Group, World Health Organization) working together to provide essential health information to developing countries, yet it appears underutilized. The BMJ reported in February 2003 that only 100 of the 438 participating institutions were using HINARI on a regular basis.[3] If their upcoming evaluation shows problems with that uptake they may wish to look to Ptolemy for ways of building their grassroots connections. HINARI is a pipeline, but it has yet to be connected to the taps. Ptolemy is a garden hose compared to the HINARI pipeline, but it actually reaches the end-users, and evaluates how they use the information it provides.

Consider the Shaughnessy equation: "The usefulness of any source of information is equal to its relevance multiplied by its validity, divided by the work required to extract the information."[4] Surgeons seldom have an afternoon to spend digging in the library and depend on after-dinner time to read the literature. There is no reason to believe that colleagues on the front line of health care in Africa are any different. Most use Ptolemy from home, not only because evening is when surgeons do their reading and research but also because they seem to have much faster Internet access from home than from work.

It is important to see that Ptolemy is not competition to HINARI, but an alternative means to the same end. Where HINARI's strengths are at the institutional level, Ptolemy's are its grassroots connections with a small group of highly influential African surgeons, the way it engages them in the evaluation enterprise, and the way it uses existing university library resources. The purpose of this project is not just to provide information but also to evaluate how it contributes to building the communities of medical curiosity needed in developing countries to solve the health problems that affect billions.

The incremental cost of providing access to our research affiliates in developing countries is negligible to both publishers and universities as it utilizes preexisting resources. As any company is concerned about both its corporate image and its bottom line, publishers may want to think about whether in the long run, it will be less expensive and more effective to build a new international library with Geneva-based bureaucracy (HINARI) or to allow existing libraries like the University of Toronto to develop Ptolemy-like projects. It is plausible that the overall costs of the latter approach will be much lower and the benefits more direct. There is no capital cost, no loss in revenue, and the impact of their journals will improve. The diversity of information needs in developing countries argues for a multiplicity of strategies to address them. Ptolemy is producing data that will be very helpful to all involved in the enterprise of bringing electronic health information to doctors in developing countries.

---

[3] R. Smith. 2003. "Closing the digital divide," *British Medical Journal*, 238.

[4] F. Godlee, R. Smith, and D. Goldman. 1999. "Clinical evidence," *British Medical Journal*, 318:1570-1.

FUTURE DIRECTIONS

We regard Ptolemy as an experiment in providing access to the medical literature and evaluating how the information is used. The next stage will involve electronic "hit" analysis to provide a clearer view of which resources are most useful to our African colleagues. Ptolemy functions to fortify a fragile emerging research community of African surgeons and we are now engaging them in a Delphi process to identify priorities for surgical development in East Africa. Our participants are, after all, the experts and so we are building the Ptolemy group into an interactive research community focused on finding African solutions to African problems. Ptolemy provides a readily reproducible model for the dissemination of the medical literature and research community building in developing regions. We are working to insert a clause in all the library's agreements with publishers acknowledging that a certain proportion of library accounts will be made available to the university's research affiliates in poor countries. We seek to persuade other universities to build similar partnerships with colleagues in developing countries. It would be particularly helpful to link African HIV researchers in a similar type of project with their counterparts at a major northern university.

13

Health Information for Disaster Preparedness in Latin America

Jean Luc Poncelet
Pan American Health Organization, United States

Latin America is high risk in terms of disaster. Natural disasters include volcanic eruptions, such as Armero with 10,000 deaths in Columbia; earthquakes in Mexico and El Salvador; El Niño; and hurricanes, such as Hurricane Mitch. In October 1998 Hurricane Mitch impacted Guatemala, Belize, and part of Mexico, leaving 9,000 dead and 13,000 injured. These are large numbers for small countries that have less than 4 million inhabitants. The hurricane displaced 3 million people and caused $5 billion in damages.

Timely information is very important for disaster management. This information, however, relies mostly on a multitude of scientific groups that unfortunately do not communicate frequently among themselves. For example, Juarez hospital collapsed in a major earthquake in Mexico City in 1985, and was unable to function. The earthquake killed patients, surgeons, nurses, and the administrators who are needed to ensure that another collapse does not occur. Hospital construction needs input not only from architects, engineers, and builders but also from physicians, meteorologists, and volcano and earthquake experts.

Many specialties should work together; however the scientific community is working on a very isolated basis on very specific topics. Poor disaster management is in part due to inept collaboration. Thus, managing natural disasters is managing information. Reliable information is the most valued commodity before and after a disaster, because it allows us to save lives as efficiently as possible.

DISASTER INFORMATION IN LATIN AMERICA

The three main categories of users of disaster information in the Americas are academics and scientists, policy and decision makers, and disaster and risk managers. The disaster and risk managers influence the decision makers in taking preventive measures in a country.

The first issue that we have to deal with is the wealth of information of which little is accessible. In addition, the available information is not always applicable. Professionals in developing countries have a wealth of information that could be applicable to neighboring countries, but very little has been written down. Even when written, this information is often not circulated or when circulated often is not in a format that is useful for decision makers.

The availability of information is increasing, but often this information is of relatively poor quality and not always applicable to the developing country's context. The best available written information in developing countries at the present time has not been peer reviewed. Researchers in developing countries prefer a non-peer-reviewed article written in their native language that is locally relevant to a peer-reviewed research article written

in a language they cannot easily read with conclusions that are not applicable to their country. While peer-reviewed publications are the most useful, less effort can be made in collecting and distributing them. Researchers in developing countries look forward to having better information and more scientific information, but they have to make use of what they have.

Little by little, limited access is becoming a myth. There is increasing access to the Internet in the Americas. The people who already have access to the Internet usually are extremely active, have a large influence on the community, and can communicate information. They are absolutely effective in their environment. It is the rest of the public that we have to work with.

THE REGIONAL DISASTER INTERNET NETWORK

Hurricane Mitch was the first Internet-dependent disaster in the Americas. It was the first time that a disaster was so influenced by communication through the Internet. Some information was reliable, some was not. The Internet was a powerful tool for coordination and provided extremely useful information for decision making. It was also a fertile ground for rumor. As a result, the Regional Disaster Internet Network was created to facilitate access to relevant and accurate information for decision makers.

Three organizations are working together on this project—the Pan American Health Organization (PAHO), the U.S. National Library of Medicine (NLM), and the Regional Disaster Information Center (CRID). PAHO is the regional office of the World Health Organization; it is 100 years old and has 2,500 staff members working throughout the Americas and the Caribbean. It has many contacts in all of the Latin American communities. That staff is not based in Washington; rather it is a very decentralized organization with approximately 2,000 people in the field. It promotes technical cooperation and knowledge sharing between countries.

NLM is a world reference health information center and offers free access to health information, including PubMed Central, the most consulted medical database on the Internet. NLM assists countries in obtaining access to the Internet and sharing expertise. That assistance is provided mostly in the United States, but NLM has worked outside of the United States in collaboration with PAHO.

CRID is a nongovernmental organization that was established in Costa Rica in the 1990s. It was originally funded by PAHO, but now is self-sustaining. CRID has 14,000 publications on disaster-related topics, mostly non-peer-reviewed. Collecting these publications on very specific health and disaster topics has been a major challenge. Quality is relatively good; it is not the highest quality in existence, but it is information people are looking for.

CRID has an average of 6,000 end users on a regular basis. The center is working with the support of many organizations and is multidisciplinary. It is the main source of documents in Spanish. At the present time most of the documents are based in Costa Rica for very practical reasons, but are being shipped to other countries. CRID also provides copies to users who do not have access to the main library in their country, which is likely located in the capital cities making it difficult for these users to access a hard copy.

CRID is also linked to a number of other Internet and library services and the different services are linked; one advantage of CRID is the ability to move directly from one database to another. For example, CRID is linked to LILACS, a CD-ROM that is distributed throughout the region. It is read by physicians and public health researchers, and hence is a system that can reach a large variety of people.

The result of the Regional Disaster Information Network is to increase the multisectorial access and use of the Internet. The use of the Internet by these disaster managers and health documentation specialists is increasing. It has been possible to set up six documentation centers in four different countries, increasing the coverage and number of fully digitized documents and making the Internet accessible to the population in the Americas. This has increased circulation dramatically and improved the document selection process, as well as the thesaurus. There was no thesaurus in disaster management until 10 years ago. Today's thesaurus is the result of a significant effort in collaboration by a number of organizations.

The collaboration between NLM and PAHO has been a very interesting coordination effort. The partnership created a synergy between NLM's good information technology capacity and PAHO's knowledge about risk management and the regional network. There is also high motivation from the participating countries. That combination has made this project a great success.

CONCLUSION

Scientific information on disasters exists in developing countries, but there is very little access to that information. The most directly applicable scientific information is mostly unpublished and found within either that developing country or a neighbor. The project demonstrates that a simple arrangement between good partners can make information available to a large number of lower-income professionals in poor countries.

The active dissemination of good quality but free information is the main asset of the Regional Disaster Information Network. It is probably the most cost-effective disaster reduction activity that exists in the region. Providing technical information to people who are interested in the topic, especially if there is political will, has a huge impact in the countries. It also empowers nationals. The dissemination of information is a powerful means to reduce the gap between rich and poor by making data accessible, and is probably the best means to reduce the impact of a disaster.

14

Bioline International and the *Journal of Postgraduate Medicine*: A Collaborative Model of Open-Access Publishing

D. K. Sahu and Leslie Chan
Journal of Postgraduate Medicine, India, and Bioline International, Canada

INTRODUCTION TO BIOLINE INTERNATIONAL

Bioline International[1] is based at the University of Toronto and has partners from all parts of the world. The goal of Bioline International, which started in 1993, is to improve the visibility, accessibility, and availability of research material from developing countries. Much of this research material is of high quality and is especially important to people in developing regions as well as to people in the North. Our goal is to make some of this science visible and become part of the global knowledge base. We also try to create low-cost models of collaboration for scholarly publishing and have been working with like-minded partners for the last 10 years.

We have been documenting the processes that work and those that do not work to share with those who are interested in replicating similar types of projects. Ultimately we want a process that is portable, meaning that anyone who wants to pick up where we started could easily adopt the same process and software infrastructures and do the projects on their own. One of the goals of this project is ultimately to eliminate ourselves so the journal partners can become self-sufficient and not dependent on Bioline. At the same time we are constantly thinking about technology transfer issues and human resource development.

Many initiatives that focus on providing information, such as the Health InterNetwork Access to Research Initiative (HINARI), are concerned with information flows from the North to the South. While that is very important, we must remember that much of the information created in the South is also important to the North as well as to the South, particularly in areas such as biodiversity and tropical and infectious diseases.

Bioline International has several partners in the project. Bioline's main partner is the Centro de Referência em Informação Ambiental (CRIA), or the Reference Center for Environmental Information. The Bioline server is located at CRIA in Campinas, Brazil. This is an example that any place with a good Internet connection can be the center from which information is disseminated to the rest of the world. The software is entirely open source and is based on Linux and Apache. The Bioline International Web site has online journals from different parts of the world. Bioline primarily helps journals that have only print publication with small circulations in their parts of the world. Bioline digitizes these journals and posts them online in a database. The University of Toronto Library, another partner of Bioline, provides technology support.

[1]See http://www.bioline.org.br.

There are many journals available on the Web site. Some of these journals have entirely free access. Others are not open access, meaning that their publishers wish to recruit subscriptions for access. Bioline has been experimenting with various models for providing access to the information, in accordance with our partner publishers' desires. Bioline is trying to educate publishers about the importance of open access; one of the good examples of this work is the *Journal of Postgraduate Medicine.*

THE *JOURNAL OF POSTGRADUATE MEDICINE*

The *Journal of Postgraduate Medicine* (JPM) [2] is a quarterly publication started in 1955. It is a publication of the Staff Society of the Seth G. S. Medical College and K. E. M. Hospital in India and covers specialties from basic and clinical sciences.

As we all know, journals from developing countries are poorly represented in international bibliographic databases. This, along with low print circulation, adds to the poor visibility of research published in journals from developing countries. This in turn leads to lower citations for the published articles, so that journals with a low impact factor continue to remain less subscribed and poorly visible.

One of the most important problems with journals from developing countries is the poor visibility of the published articles. Electronic publishing offers a solution for increasing the visibility. However, the financial and technical issues of electronic publishing make it difficult for journals of developing countries, which are usually not supported by commercial publishers, to go online. Even if a journal succeeds in doing so, a single journal site fails to attract visitors, because it alone can provide only a few hundred articles. There is a lack of interactivity and hyperlinking, which makes it less attractive and less useful for visitors.

A collaboration with Bioline, or any other established portal, offers shared resources and technology that is already tested. Bioline acts for journals from developing countries as PubMed Central does for countries of the North by eliminating the technical difficulties. The crosslinking and hyperlinking associated with Bioline make full use of the publishing potential. Journals get the advantage of established reputation and greater visibility by linking to the Bioline site.

Apart from providing the full-text content to Bioline, the JPM has also helped to promote collaboration with the Bioline site. It provides linking from PubMed, the U.S. National Library of Medicine's largest database in the form of a LinkOut. JPM may also help in technology transfer to Bioline by providing tools that we developed, such as automated reference linking. JPM also encourages other journals from developing countries to join Bioline and provide open access. There is a symbiotic relationship between the journal and Bioline.

What has been achieved with JPM's collaboration and Bioline's provision of open access to the journal? The most important achievement was the archiving of JPM's full text. This, along with the JPM Web site, has helped to increase the visibility and readership of the journal, which has had a direct impact on the number of citations of the published articles and on article submission from around the world. The journal has also gained an international reputation and now is included in a large number of bibliographic databases. Since providing open access two years ago, the number of submissions per month has increased threefold. Currently more than 40 percent of submissions are from other countries.

This means that the journal is becoming a popular publishing medium for scientists from other developing countries. How is this possible? The circle of accessibility (see Figure 14.1) has played an important role in increasing the popularity and visibility of the journal. The journal's content is available in full text and is linked to a large number of resources to increase its visibility.

JPM encountered some problems in developing its partnership with Bioline. Initially JPM had difficulty transferring huge amounts of data from its site to Bioline's technical team, but with use of file transfer protocol this has been solved. In addition, JPM is not able to update on a regular basis because the site is maintained by Bioline's technical team. If such collaboration breaks, a journal that is associated with Bioline should be able to continue independently.

---

[2] See http://www.jpgmonline.com.

LESSONS LEARNED

Bioline's collaboration with JPM is a concrete illustration of the direct impact of open access. This is a successful lesson that Bioline is teaching to other journal publishers in developing countries. The main problem in journal publishing in such countries is not poor circulation but poor submission. Poor submission fuels a vicious cycle. Authors are afraid to submit to journals that have poor circulation because they want their article to be read. If authors do not submit, journals do not come out regularly. If publication is irregular, people do not want to subscribe. The *Journal of Postgraduate Medicine* is able to publish six issues a year because of the high submission rate. This is a very good means of attracting authors, increasing the readership, and possibly increasing the readership rate for the journal.

Many publishers are concerned that providing open access will hurt the publication; however, these models often improve the journal in many aspects, as the JPM example illustrates.

Someone must pay for the production of the journal. The question is, who pays? Many developing country journals are funded by international funding agencies of various levels. We should analyze how these agencies are funding these journals rather than put too much constraint on the publishers who have to charge a paid subscription to cover the costs. There has to be a way for publishers to recover the investment of the funding agencies, not just from the subscription costs but from other means as well.

We should consider new sustainability models. In addition to learning more about the open-access models, other plans for Bioline include expanding to include other journals, such as the *East and Central African Journal of Surgery*. We recently signed an agreement with the *African Journal of Reproductive Health*, which provides important scientific information on material related to reproductive issues and health issues for women in Africa.

We hope to promote Bioline more to library consortia. The more these consortia are linked up to the Bioline database, the more visibility and hits our publishers will get. Hopefully this will feed back to that circle of accessibility illustrated in Figure 14.1.

Bioline is moving in the direction of open access for all the journals on its system. Providing closed access is very expensive. Most of Bioline's time in terms of administrative costs is providing lock-up for the journals that want to charge for access. For example, single payment for an article is $8, and we figure that it actually costs Bioline more to collect the $8 than it is worth.

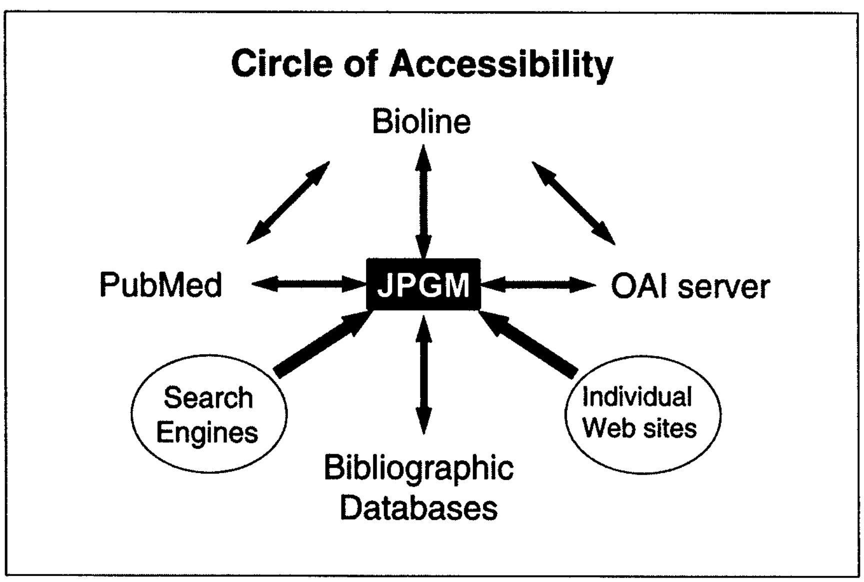

FIGURE 14.1 Circle of accessibility.

We are in the process of implementing extensible markup language for full text for the articles, which would allow much more sophisticated citation linking and metadata retrieval. Currently Bioline's abstracts are Open Archives Initiative (OAI) compliant. We are also archiving the open-access journals on a Bioline Eprints server.[3] The OAI and Eprints servers initiatives are trying to improve the interoperability of databases that are scattered across the Internet. This again is a tool to increase the visibility and long-term availability of scientific information.

We are hoping to become multilingual to accommodate the different languages in which the journals from Africa are published.

We can always improve the user interface, but more importantly Bioline wants to track how users actually use the information. We may have many hits and downloads on the Web site, but what are the users doing with the information they have downloaded? The Ptolemy Project gives us a window on that particular aspect of user behavior.[4] Bioline must set up multiple mirror sites so that there will be a site for the local journals as more partners are created in Africa and India. This site will provide even faster access and eventually allow them to take up the operation on their own.

In conclusion, we are heading in the direction of self-sustainability for these journals, however we should find new models for sustaining them. UNESCO, ICSU, CODATA, and the Open Society Institute should be thinking about open access as part of their overall strategies for funding journals from developing regions of the world.

[3]See http://bioline.utsc.utoronto.ca.

[4]See Chapter 12 of these *Proceedings*, "The Ptolemy Project: Delivering Electronic Health Information in East Africa," by Massey Beveridge.

SESSION 3: DATA AND INFORMATION IN THE ENVIRONMENTAL SECTOR

15

Introductory Remarks by Session Chair

Farouk El-Baz
Boston University, United States

Worldwide emphasis on fast-paced economic development necessitates attention to the environmental consequences. Improving and sustaining the environment are prerequisites for proper development in the long term; therefore, monitoring changes in the environment becomes an essential tool for sustainable development. Digital data and information pertaining to the environment are an integral part of the discussion of open-access and public-domain information. Satellite image data from numerous sources (on the atmosphere, the land surface, and coastal zones) form a basic component of the requirements. Two or more digital images of the same object or phenomenon can be superimposed, using widely available tools of information technology hardware and software, to instantly display the change from one time to another. Such data should be made available as widely as possible to the international scientific community. Three distinguished contributors to this session will cover the following aspects of this endeavor.

The first contribution is by Mukund Rao of the Indian Space Research Organization. His presentation emphasizes that the use of geospatial information is limited only by the imagination, citing examples, including natural resources development, disaster management, land use assessment, and environmental restoration. The second presentation, by Peter Weiss of the U.S. National Weather Service, deals with meteorological data. Open and unrestricted access to meteorological data is provided in the United States free of charge. However, in Europe these data are considered revenue generating in the short term, which limits their availability and use. The third contribution, by Liu Chuang of the Chinese Academy of Sciences, presents a case study illustrating that the current access models for information in China are changing and moving toward free and open availability.

There is no question that access to information can lead to a better understanding of Earth and its environment. Efforts should be made to constantly assure the availability and ease of exchange of data among researchers worldwide.

16

Geospatial Information for Development

Mukund Rao
Indian Space Research Organization

In May 1983 a few eminent Indian scientists visualized an institutional framework of the National Natural Resources Management System that would enable the utilization of spatial images from Indian satellites to support national development activities. Spatial information was visualized from the beginning as a resource that could be used for a number of applications by a number of users in the government, the private sector, and academia, and by the public. Spatial information could be used by the government for land management, property management, water resources management, environmental assessment, and so on. The private sector would need it for business applications, development, and marketing services. Academia would be a major user of the spatial information, using the information for research, such as global change research and climate modeling. Private citizens would also be major users of spatial data. The spectrum of users of spatial data is quite wide and today in India it is a resource that can serve all these users.

Satellite-based observations across India are a major source of spatial information for developmental activities. India has satellite images from 1 km resolution, which provides images every half an hour, to 1 m resolution. The satellite images should be converted to spatial information, such as land use maps, groundwater maps, and forest maps, and then these spatial data sets should be linked with nonspatial data, such as human census data and village development data, to make them more useful. The position information from the Global Positioning System satellites can give the precise location information that is required for high-resolution satellite image analysis today. Satellite images can provide cartographic-quality maps, terrain models, models of topography, cadastral maps, and support a whole systematic mapping process.

USE OF GEOSPATIAL INFORMATION FOR DEVELOPMENT IN INDIA

There are many examples of how satellite images and related spatial data and information are used in India for development. Spatial information forms the basic unit for supporting the area development process and planning development. Indian remote-sensing satellite data for the entire country have been used to map the wasteland areas. A digital database containing the basic wasteland information is used to plan the use of land, and such plans for land management are now available.

Spatial information is also used to find sources of drinking water in India. A very exhaustive map is prepared using satellite images and other information to assist villages in finding drinking water within a 1 km radius that

can be tapped. This information is regularly used by district administrators and decision makers to solve many of the drinking-water problems that they have in their areas.

Spatial information is also useful for crop production estimations. Every year satellite images are used in conjunction with meteorological observations and market information to provide four estimates of crop production in a crop growth cycle. This is done for wheat, paddy, groundnut, jute, cotton, and a variety of other crops for the entire country. This information, which is based on satellite images and integrated models, is used by decision makers to plan the stocking and movement of food grains and also on decisions for food import and export. This is a national project conducted by the Ministry of Agriculture in India.

Spatial information is also used in urban development. Annual monitoring of cities is done using satellite images, which are mapped and studied to find out how a city is growing. A digital database is created and the exact location for the future growth is modeled to determine the type of development a city should adopt.

Disaster management is an area where space images play a very major role. For example, the Indian Remote Sensing Satellite observed the flooding in the state of West Bengal. The flood inundation information was abstracted from the satellite images and superimposed in digital format onto a map. Such statistics as extent of inundation, villages affected, and areas damaged are passed to the district administrator who decides the type of rescue operation, damage assessment, and measures needed to see that these disasters are mitigated.

Drought occurs quite frequently in India. There is a program that uses satellite images to generate the weekly normalized difference vegetation indices (NDVI) during the summer season. The NDVI is used as a reference and compared to an estimate to find out whether they differ. The NDVI is modeled across time to discover the onset of a drought. This operational program, which is done using satellite images and related information, is used by administrators to decide whether any mitigation measures should be taken for the drought.

India has a variety of biodiversity hot spots, and a unique application focuses on biodiversity characterization. This project involves using satellite images to map these hot spots, as well as forest density and forest types. Similarly, satellite-image-based modeling is done for a species assessment at the ground level. These maps are augmented by in situ collection of species information to generate a biodiversity index that categorizes areas as rich or poor in biodiversity. This program is done using geographic information systems (GIS) and satellite imagery.

Another application uses terrain models at the district level to determine where water conservation efforts should be directed. This prognostic terrain and land use information is generated using GIS and spatial modeling tools and is available for the district administrator to plan water-harvesting structures for that area. This information also assists district administrators with decisions about employment opportunities or subsidies.

A major program in India has been the establishment of a Natural Resources Information System in which an integrated GIS database of 22 spatial and 8 nonspatial parameters is generated as decision support for district authorities. This integrated database, apart from supporting natural resources management, is used at a district level to determine the location of such amenities as primary schools, health centers, and general facilities. Criteria have been developed to assist with narrowing down sites to determine the optimal location for these facilities.

BUILDING A NATIONAL SPATIAL DATA INFRASTRUCTURE FOR INDIA

Over the last 20 years India has generated a large archive of geospatial data and information. Until recently most of this information was in different formats, and many different agencies were working on it. The data were also in different scales. There were various obstacles limiting access to this information, including difficulty in finding the information, reproducing it, and using it in GIS to support development activities. The government is now establishing a National Spatial Data Infrastructure (NSDI) to standardize formats, standardize archival of spatial data, and make them easier to access. The underlying idea is to adopt an open-access approach and view the spatial information of India as a national resource.

Many databases have been generated using the satellite images, as well as the related information derived from those images. This information supports national, state, and district development projects. NSDI provides access to organized spatial data and information. The goal is that the infrastructure is used for a sustained economic growth.

Multiple servers are being established in different agencies to support the NSDI. For example, the Soil Department has digitized its data and put up a soil server that has all the soil information. The Environment and the Forest Departments have put all the forest information and the satellite images on separate servers and so on. Thus, NSDI provides a single gateway that allows users to go to these servers and extract the information they need.

Access rights and users must be defined for the NSDI to be operational. The public access will be mainly in a preview mode. The generator of information will have the full right of access. Other policy issues involving costs, regulation of access, and classification of information, for example, are currently being defined.

A server that contains the metadata of the spatial data is available. Through these metadata one can find out the information that is available in different agencies for different areas. This metadata server will eventually be linked to the individual digital GIS databases, which will reside on different servers. The user will then be able to go to the metadata and access the information through the total network that will be established for the NSDI.

In addition, a remote-sensing data policy has already been approved by the Government of India. This policy clearly states that the satellite images are seen as a public good. Through this policy the government will ensure the continuous availability of the Indian remote-sensing satellite images through a series of future missions.

Satellite images of up to 5 m resolution are available on a nondiscriminatory open-access basis, but for the high-resolution images that are 1 m, access is regulated by authentication of the user and the use of the application. This process ensures that a right to add value is given to the user, while the original copyright of all the satellite images rests with the responsible national agency.

We are slowly moving toward consolidating the issues of a spatial information policy. We are trying to position spatial information, which includes not just the satellite images but also a variety of other information from photographic service, cartographic service, the GPS, and ground tools service as a societal good. This is especially important if these data and information are going to be useful for development. One constraint that surfaces in providing complete open access is the strategic value of high-resolution satellite imagery and large-scale map information from a national security perspective; this type of information would not reside on the NSDI.

A major concern of most of the agencies is their rights as the information producer. How does the NSDI defend the rights of the information producer? Other issues involve copyright, which is complicated in the digital domain and includes licensing, as well as technological protection measures.

The crux of the whole policy is the right to access. Who has the right to access all the spatial data and what are their rights? They should be very clearly documented and known; transparent policy is critical. Different user groups should be granted different rights. The public should have certain rights; an individual should be able to access data that are required by the public. The government should have different types of rights because they are the actual owners of the information. The private sector should be given a different type of right for accessing the information, for operating the database, and producing it for development. Academia, for research purposes, should have yet another type of rights. The process of defining these categories of users and their rights is ongoing.

There is a major debate concerning the right to add value to the data and information and how that right is defined. What is value addition? If someone changes just some attributes of the data do they become value-added? The scope by which one defines the term "value-added," especially in the spatial domain, needs more clarity. Other issues involve the marketing of value-added data, as well as the rights of the producers of the original raw data. Map information is treated as value-added because the map is a derivative of the process. The development of an NSDI engenders even more questions. For example, could some of the spatial data be used against individuals and society? There are also privacy issues, both individual and societal.

NSDI must also consider cost models. One model, which charges the cost of reproduction and dissemination, would be applied mainly to most of the government-owned information because the government has the obligation to society to generate the right information and make it available. Of course, the commercial, or total cost recovery model, is another to consider, particularly when information is in the commercial domain.

It is anticipated that the NSDI strategies and guidelines will be released in the near future. After discussion within the government, it will then be made public how the NSDI will actually be accessible or available to different groups of users.

17

Borders in Cyberspace: Conflicting Government Information Policies and Their Economic Impacts

Peter Weiss
U.S. National Weather Service

What does "borders in cyberspace"[1] mean? Everybody knows that there are no borders in cyberspace. Cyberspace has completely eradicated all national borders. Anybody, anywhere can get information and do anything with it. Right? Not quite!

Let us take the true case of an atmospheric researcher in India fascinated by a problem that affects over a billion people—monsoons. The prediction, with some level of skill, of the onset, duration, and severity of the cyclical monsoons would greatly benefit society. Researchers have developed a number of analytical techniques using computers that allow some understanding of the coupled earth, ocean, and atmosphere system. To develop a method to predict monsoons the researcher develops a hypothesis that integrates American and European daily global atmospheric model output results for the past 30 years into a large global database. Those observations are compared to actual observations of what the monsoons have done over the past 30 years to tease out some trends and evidence that might allow the prediction of monsoons, which could benefit literally billions of people. This is pure basic research; no one yet knows whether it is possible to predict the monsoons.

To conduct this basic research the researcher contacted the U.S. National Climatic Data Center and got access to 30 years of model output data essentially for free. The Indian researcher then contacted the European Center for Medium-Range Weather Forecasting in England and requested data for the same approximate 30-year time period. However, the European Center as an E.U.-chartered organization is required to charge the researcher for access to these data. The researcher did not have adequate funding and now has to attempt to do this research with the U.S. data only.

The Indian researcher just hit a border in cyberspace, a border between two policy domains. Geographic borders are decreasing in importance, while borders between "policy domains" (e.g., the rules and practices associated with the access to and ability to use scientific, environmental, statistical, and other data) are becoming increasingly important.

[1]For additional information, see Peter Weiss. 2002. "Borders in Cyberspace: Conflicting Public Sector Information Policies and Their Economic Impacts—Summary Report," at http://weather.gov/sp/Borders_report.pdf.

U.S. PUBLIC INFORMATION POLICY

It seems to be a well-kept secret in Washington that the U.S. policy of open and unrestricted access to taxpayer-funded government-generated public information is not based solely on abstract notions of democratic principles, such as open government and transparency. Rather it is also based on an economic understanding that the information generated by the single largest information generator—the government—is an important input, just like gas, coal, or water, to the economic process. The United States holds as a matter of both economic and political principle that "government information is a valuable national resource, and . . . the economic benefits to society are maximized when government information is available in a timely and equitable manner to all."[2]

The roots of this policy run deep. The sources of U.S. information policy include the 1976 Copyright Act, the Freedom of Information Act (FOIA), the Paperwork Reduction Act, the Electronic FOIA Amendments of 1996, and the Office of Management and Budget Circular No. A-130. Information generating agencies of the U.S. federal government, including the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration, the U.S. Geological Survey, the Department of Agriculture, the National Institutes of Health, the Census Bureau, the Bureau of Labor Statistics, and all others, follow the same information policy, which is to actively disseminate all taxpayer-funded public information without any restrictions or conditions and without the assertion of copyright or database protection regimes.

The United States does not have a *sui generis* database protection regime nor is it likely to have any such regime in the near future. U.S. government agencies are forbidden to charge more than the cost of dissemination for the information and they are urged to take advantage of private and academic channels of information dissemination. Agencies are encouraged to use the best available technology, including the Internet and the World Wide Web.

Taxpayer-funded government information is an important economic input. The database and information retrieval industries in the United States are large and have grown exponentially since the beginning of the Internet revolution. They are dependent to a great extent on free, unrestricted, taxpayer-funded government information, which covers everything from economics to statistics to agriculture to the weather.

Approximately one-third of the U.S. economy, and likely most nations' economies, particularly in developing countries, are weather and climate related. Every sector of the American economy has some sensitivity to the weather, and firms that cater to that sensitivity have developed in the United States. Firms cater to other sensitivities as well. For example, the United States has a very robust geographic information industry that one does not see in many other countries.

Weather and climate information, and utilization and exploitation thereof, is a partnership in the United States. The government works very closely with the research community. The advances in critical engineering and analytical techniques that the researchers in India are seeking to use in their research on monsoons are dependent upon both academic and governmental support.

At the same time we have in the United States a number of interesting ventures involving the private sector and the media. For example, the U.S. company The Weather Channel has no counterpart in Europe. I suspect that "borders in cyberspace" is one of the reasons.

DIFFERENT MODELS FOR FUNDING PUBLIC-SECTOR INFORMATION

There is an interesting bureaucratic philosophical difference between information policy in the United States and in Europe (see Figure 17.1). The U.S. model says that if the mission of the Environmental Protection Agency, National Oceanic and Atmospheric Administration, National Institutes of Health, or U.S. Geological Survey is to create information, then that information should be made available to the users, who are taxpayers and contributed to the cost of generating the information in the first place, to do with as they will. The users in turn will generate jobs, income, new business, and new industry using the information.

---

[2]Office of Management and Budget Circular. 1996. Circular No. A-130, "Management of Federal Information Resources," 61 *Federal Register* 6428, February 20, at http://www.whitehouse.gov/OMB/circulars/a130/a130.html.

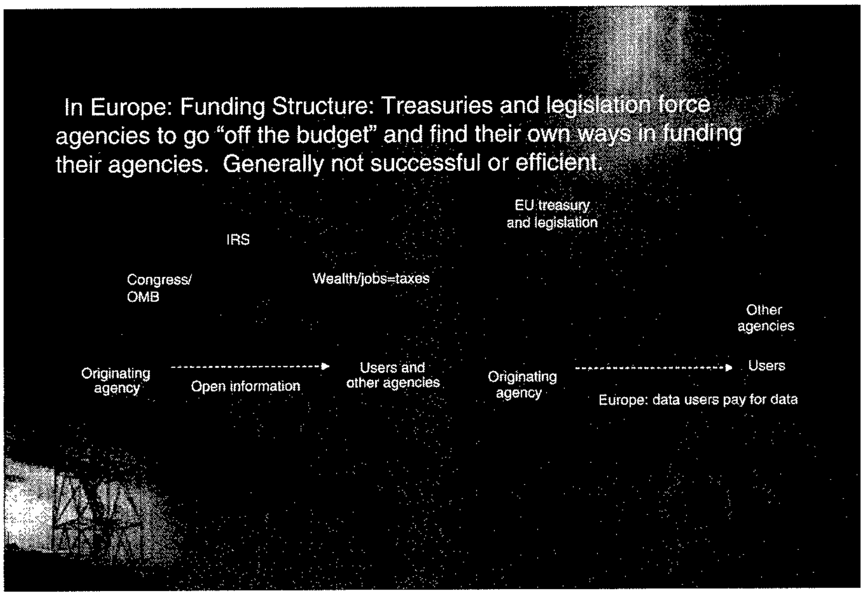

FIGURE 17.1 Different funding models for public-sector information.

This differs from the European concept of public-sector information. In Europe the funding structure is different from that of the United States. In general, European treasuries and legislation force agencies to go "off budget" and find their own funding. Often agencies charge users for use of their data and information. These users can be other government agencies and academics, and are not limited to "commercial" users. This leads to "government commercialization," which occurs when government agencies charge the public for information services that were previously considered "public goods" and were financed by tax revenue. This trend is also known as cost recovery. Government commercialization should not be confused with privatization, where functions that are not inherently governmental (e.g., utilities, telephone services, airlines) are transferred to the private sector.

When thinking about the economics of information, it is important to remember that information is not a normal good in the economic sense. Basic laws of supply and demand work differently in the information world. Information is nonrival and nonexclusive; that is, information is a public good. While information has high initial fixed costs of generation or collection, it has generally low reproduction costs.

Only in the last couple of years has the economic community started to study these very issues. It should be noted that the economies of the European Union and the United States, for the sake of this discussion, are approximately the same size in terms of gross domestic product. The United States spends approximately twice as much of taxpayer dollars on the development of public sector government-funded information as the European Union. The return on that investment, in terms of commercial growth, job growth, and taxes paid to the treasury per dollar, is approximately five times larger in the United States than in Europe.

Why is that? My hypothesis is that the United States actively encourages the dissemination and use of that information in commerce, academia, and the media, while in Europe they do not. PIRA International recently conducted a study commissioned by the European Commission's Information Society that examined the potential of public-sector information in Europe.[3] They found that charging for public-sector information might be counter-

[3]PIRA International. 2000. *Commercial Exploitation of Europe's Public Sector Information,* Final Report for the European Commission, Directorate General for the Information Society, PIRA International, LTD, University of East Anglia and KnowledgeView, Ltd.

productive, even from the short-term perspective of raising direct revenue for government agencies. They also noted that the fledgling E.U. information market would not even have to double in size for the government to recoup in extra tax receipts that they would lose by ceasing to charge for public-sector information.

As previously mentioned, one-third of the U.S. economy is weather sensitive. For example, an energy firm selling natural gas makes more money during a cold winter; during a warm winter they make less. A beach resort in Florida makes more money when the weather is warm and sunny; if it is cold and rainy, they make less. A ski resort in the mountains makes more money when it is clear, cold, and snowy; when it is overcast without snow, the ski resort makes less. In the United States someone interested in hedging their weather risk can purchase financial instruments from various companies.

The United States has a booming weather risk management industry compared with the smaller weather risk management industry in Europe, much of which is in the Netherlands. The Netherlands has adopted an open data policy for its meteorological information, both present observations and their historical data, but France, Germany, and Great Britain have not. As such, not only is there an economic disparity between the United States and Europe but there also exists an economic disparity within Europe based on data and information policies. This is another border in cyberspace.

In the August 2002 issue of the German popular geography and natural history magazine *Geo*, there is an article that discusses information policy. *Geo* claims that the commercial meteorology industry in the United States is 10 times the size of the counterpart industry in the European Union due to differing data policies regarding information as an economic driver.

A researcher in Great Britain who was developing a major international global database on weather and climate approached the U.S. National Climatic Data Center for all the U.S. historic weather observations from 1948 to the present. He received 15 gigabytes of data, for which he paid $4,200. The researcher made the same request of the German Weather Service, which quoted him 1.5 million DM (over a half-million U.S. dollars) for the historical data of all of Germany and 4,000 DM for the historical record of only one weather station. Because the researcher could not afford the data, he did not buy them. The German Weather Service did not sell the data, and thus did not make any money. The problem lies in the bureaucratic tendency to use government assets to attempt to raise money but without the incentive structure of true private-sector firms. Government commercialization is extremely tempting to bureaucrats, particularly to officials in national treasuries.

The United States has attempted on numerous occasions, at both the federal and state levels, to create user charge regimes for public sector data and information. Those attempts have failed because of the basic economics of information, public goods, nonexclusivity, and the inability of governments to act truly as businesses.

Europeans are starting to figure this out. European treasuries are still pushing European agencies quite strongly to engage in cost recovery. Many European agencies get transfer payments from other government agencies, and therefore consider themselves to be self-supporting. For example, 50 percent of the U.K. Meteorological Office's revenue comes from the Defence Ministry and another 30 percent from other government agencies. This is essentially an accounting transfer from one government account to another government account.

The German Weather Service recovers only approximately 1 percent of its costs from data sales. It was embarrassed when the federal audit office realized that not only did it not account for its costs adequately but it also created a price structure that put it in violation of German competition law when it came to dealing in the private sector. That helps explain why there are very few commercial weather services in Europe.

CONCLUSION

Cost recovery is not the best approach for maximizing the economic value of public-sector information to society as a whole, not even from the viewpoint of government finances. Prosperity effects will be maximized when data are sold at marginal cost. Direct government funding and free provision to all are favored with their contribution to national welfare maximized at the point where marginal benefits equal marginal costs. There is a growing recognition in Europe that public-sector data policy is critical to the information society. I hope the World Summit on the Information Society will discuss this topic comprehensively.

Recent trends towards more "liberal" policies still face opposition from national treasuries and entrepreneurial civil servants in charge of government commercialization initiatives, which can result in anticompetitive practices that hinder the growth of private-sector competitors. A tension thus exists. There are two European Commission directives on public-sector information and environmental information that are under consideration and relate to this concept.[4] Both directives recognize that public-sector information is economically important, both generally and in the area of the environment. These directives are grappling with the two major issues: (1) pricing policy and the cost of dissemination and (2) the downstream uses of information and the restrictions on such uses. There have been some reforms at the national level in Great Britain, the Netherlands, Sweden, Finland, and Germany.

The United States is not alone in terms of open information policies; Australia, Canada, New Zealand, and Japan have open policies as well. Some other countries have not addressed this issue, although China is currently studying its information policy.[5]

Global data sets are needed to predict large-scale phenomena, such as the monsoons studied by the Indian researcher. Other projects also highlight the importance of information access to development. For example, a weather radio network being set up in rural Africa called RANET is a pilot project in six countries that will assist farmers by providing accurate weather information.

Information is worthless unless it is provided to villages when they need it and how they need it. The World Bank also is thinking about weather risk management in a broader context, particularly in global agriculture, as illustrated by the first international weather risk management contract that was issued to ensure the Moroccan sunflower crop.

[4]Commission of the European Communities. 2000. "Proposal for a directive of the European Parliament and of the Council on the re-use and commercial exploitation of public sector documents," COM207, Brussels, July 5; and Council of the European Union. 2002. "Common position adopted by the Council on 28 January 2002 with a view to the adoption of a Directive of the European Parliament and of the Council on public access to environmental information and repealing Council Directive 90/313/EEC," 11878/1/01 Rev 1, Brussels, January 29.

[5]See Chapter 18 of these *Proceedings*, "Recent Developments in Environmental Data Access Policies in the People's Republic of China," by Liu Chuang for additional information.

18

Recent Developments in Environmental Data Access Policies in the Peoples' Republic of China

Liu Chuang
Chinese Academy of Sciences

I would like to discuss the opportunity that exists for developing countries in open access to scientific data, using the recent developments in environmental data access policies in China as an example. There is a new milestone in scientific development in China. A new strategy for data sharing was developed in 2002. This presentation will discuss the transformation toward this new strategy, environmental data policy reform in this process, the challenges this strategy presents to China, and some lessons learned thus far.

CHINA'S NEW SCIENCE STRATEGY

China established a new national science strategy in 2002 "to promote the implementation of a national strategy in education and science and enhance the innovative ability of science and technology of the country." The aim of this strategy is to enhance open access to digital scientific resources as well as infrastructure.

The impetus for developing these open-access principles is found in China's long history, which has been well documented in cultural records. China has more than 1,000 different kinds of yearbooks that deal with topics such as economics, forestry, as well as many others. There are also 5,000 to 6,000 major databases that have been developed and are updated regularly. In addition, different agencies house different data, so there are also thousands of data nodes in China. Yet, despite the abundance of data resources, many scientists, students, and academics have not been able to access these data.

Before 1970 all of the data records in China were paper-based but freely available. Users were able to access these data from the government and the libraries. Beginning in 1978 two categories for scientific data were created. One was confidential data. The other was for public use, but with the advent of the digital age data were now available for a fee. This market-oriented approach to data access was a big problem. In 1994 Professor Sun Shu, among others, submitted a complaint letter to the government asking for policy reform.

Chinese policy makers recognized the need to enhance data sharing and slowly began to consider how to make data more openly accessible. This was fueled in part by China's entry into the World Trade Organization (WTO). Membership in the WTO was very important for China, making it open its doors and learn from international organizations and other countries. In 2001 the Ministry of the Information Industry of China started the Government Online Program. The Electronic Administration Program was established in 2002 by the National Planning Commission. Both of these related programs are huge endeavors.

Currently in China scientists have a voice in changing policy. Scientists have submitted several proposals to the central government, which was already considering how to change policy to form new strategies in science. In 2002 the National Council approved the proposal to start the new strategies, which would be operated by the Ministry of Science and Technology of China. In January 2003 the Ministry of Finance launched a long-term budget for the new program.

There are five major elements in the transformation to the new strategy: scientific databases; information, including journals; big equipment; biological specimens, especially for agriculture; and an information network. In addition, five sectors support the national strategy and market planning. The first is legislation, which establishes the different levels of law and regulations that govern the strategy. The second is budget, which tracks the money needed each year. The third is a new administrative approach, which is very challenging. Fourth, new technologies and standards are needed to support the strategy. Finally, a team of human resources is also required. With regard to the latter element, China is seeking international cooperation to learn from and work with other national and international organizations to help China make this transformation a success.

ENVIRONMENTAL DATA POLICY REFORM IN CHINA

Environmental data policy reform in China is a big issue. The environmental data policy was chosen as the first open-access project because of its importance. In addition, environmental data have many users. Meteorological data were the subject of a pilot project in 2002. In 2003 hydrological, seismological, and cartographical data and remote-sensing data were added.

Remotely sensed data in China include Moderate Resolution Imaging Spectroradiometer (MODIS) data from NASA. Initially everyone held their own data and did not share them with others. For example, students and professors at the agricultural university in Beijing could not obtain data. This was an important problem that illustrated that unless data are shared the government money is wasted.

In the end Chinese policy makers recognized three different categories of data and information and set up different legal frameworks for each mechanism. The first category is confidential data and information, which are protected due to serious national security concerns. The second category includes commercial data. The final category is public-domain data and information.

China has also passed a survey and mapping law along with several other regulations; one is an order of the President on access to the meteorology database of China, while other regulations concern access to the survey and mapping database, as well as to the geology, hydrology, and seismology databases.

NEW CHALLENGES FOR OPEN ACCESS TO SCIENTIFIC DATA AND INFORMATION IN CHINA

China is currently facing great challenges. One challenge is making the government agencies' data available for research and education. The U.S. Office of Management and Budget Circular A-130 data and information policy is a good example, but it simply does not work in China. This is because the budget of China's central government is not large enough, and each agency has many people working on these issues. If the agencies work together, people will lose their jobs, which is a big problem in itself.

The second challenge is dealing with the data coming from research projects. This relates to how the government balances the individual principal investigators with the various research institutes and data centers. Many researchers want to share their data, but do not have the resources to do so. In addition, to which data centers should researchers submit these data and who funds the operations of these data centers?

The issue of open access to Chinese data by international users is another challenge. Internal use is considered acceptable because the Chinese pay taxes, and should be able to use government-generated data and information for free. Outside the country, people do not pay taxes to the Chinese government, so should international users be charged a fee? Policies should be established to deal with this issue.

Several more challenges remain. One involves the role of data centers and libraries in providing open access to the data and the information. There is also a challenge in managing the data with updated technology, especially

for archiving and dissemination. Data archiving is a very critical issue in developing countries, especially in China. Another challenge is how to link the data communities with research communities, which should be discussed within the data community.

LESSONS LEARNED

China needs at least five years to complete this planned transformation. From our experience thus far we have learned that there is an opportunity for industry and developing countries in open access to scientific data. It is important to share experiences to try to understand and learn from each other. For example, the United States has a policy of full and open access to government-funded data and information. Other models that should be studied are employed in Europe. We should talk about these issues, and international organizations should assist in providing these opportunities. Keeping the communication channels open is very important.

China should seriously consider how to make this transformation a success. Scientists, of course, play a very important role because the transformation is related to new technology and knowledge. The government plays a critical role in the transformation as well. Scientists and government officials should work together. The upcoming World Summit on the Information Society in Geneva provides a good venue for such cooperation.

We also should link the data community with the research community, as previously stated. CODATA and the ICSU data community should work with the research communities.

Finally, it is important to pay attention to training and recognize the role of young scientists in the new data and information technology transformation.

SESSION 4: BASIC SCIENCES AND HIGHER EDUCATION

19

Introductory Remarks by Session Chair

Lulama Makhubela
National Research Foundation, South Africa

Institutions of higher education are not disconnected entities. They exist in a social setting and operate within a context with set values and norms specific to that social setting. Therefore issues of open access, basic sciences, and higher education cannot be addressed devoid of other factors that are intrinsically linked to the political economy and power relations in the South-North divide and of late within the growing divide between countries of the South. The questions of who owns what, who has a right to access what, and who is prepared to share what have been plaguing attempts to share resources in the various library and information resource consortia at national, regional, and international levels but even more so in developing countries.

The politics of consortia and the very complex and intricate nature of sharing knowledge in basic sciences and higher education are not new, of course. However, the digital divide between North and South countries and the financial pressures on public research institutions have made it even more complex and continue to threaten the very fabric of open access of information in the public domain, because physical access and high-level connectivity to the Internet are not the factors ensuring open access. The broader question of how much is technology a tool for development still remains.

The issues at stake involve an essentially linear process that starts with the recognition of the usefulness of information and communication technology (ICT) to implement and maintain the appropriate supporting structures (e.g., financial, technological, and political). Given this process many countries are at varying levels of implementation, with some waiting to recognize the benefits and others already significantly benefiting from ICT.

South Africa ranks quite high in countries ready for the networked world[1] and compares favorably with the first world's sophisticated ICT infrastructure. Other socioeconomic, as well as scientific and technical, world indicators show, however, that South Africa lags behind in economic development, human resource development, or readiness for the information society. The issue of access has multiple aspects and hence workable solutions will necessarily have to include all the relevant dimensions. Therefore, a distinction should be made that having high Internet connectivity and a few streets of Bloemfontein, Cape Town, Durban, Johannesburg, and Pretoria lined with gold do not necessarily reflect the abject poverty and the high levels of illiteracy still facing the critical masses in South Africa and the broader South African Development Community (SADC) region.

[1]See G. S. Kirkman et al. 2002. *The Global Information Technology Report 2001-2002*, Oxford University Press, New York.

The reconfiguration of the higher education system that occured due to the political expediency of addressing the much-needed transformation of that system resulted in the reduction of the 36 institutions of higher education to 21. The vociferous debates in the reconfiguration of higher education institutions were mainly about institutional governance, size, and shape; very few voices on open access and the public domain in digital data and information for science were heard. The critical discourse on epistemological access was silent and still remains to be heard. The need for analytical capabilities, the need to interpret data and make critical judgment of what is relevant, and the interpretive capacities in association with the historical socio-economic context of communities form a critical link in the open-access value chain.

The complexity of the issues raised above results from the dichotomy between open access and protection of intellectual rights on the one hand, and the linking of the political and economic debates on the other. These debates caution us of the move away from a rather simplistic approach to the world's problems that is devoid of an understanding of power relations in world states. The divide and the concomitant problems hindering the progress to open access and the public domain in digitized data and information science have been adequately covered. However, there has been relatively little evidence of sufficient progress in this regard. On the one hand, there is a need to share information, but on the other hand an inherent nature that does not want to share also exists.

The ICT revolution has brought an unprecedented flow of databases and information resources to our fingertips. In the library and information science world we are inundated by a number of consortia. Again in South Africa there are many higher education consortia that are bringing institutions together. However, it is not an easy task to ensure that data and information of a scientific nature are shared because of the inherent nature of human beings; we are unwilling to share in some instances. The question is, if I am rich, why should I share with the poor? That has been one of the serious obstacles facing the need for information sharing.

There are several critical issues that challenge us in pursuing basic sciences in higher education. The first challenge is the representation of our higher education values. The second is paying particular effort not only to institutional access but also to assurance that epistemological access appears as well. The third issue is collaboration, not only from an economic paradigm approach but also from a humanitarian approach that will filter down to under-resourced and under-represented communities throughout the world. The fourth issue is the compartmentalization of knowledge. Do scientists engage in the consequences of their laboratory work? In South Africa the high-profile scientist Wouter Basson invented some weapons of mass destruction, which destabilized the country and created problems of acceptance of science by society as a whole. We must deal with issues that concern higher education and compromise the very scientific nature of investigation.

The Symposium on Open Access and the Public Domain in Digital Data and Information for Science is therefore not only relevant in giving expression in the convergence of the various science domains and their complimentary nature through the advent of the ICTs but also to cement ICSU's goal of bringing together scientists worldwide in nongovernmental, international, and interdisciplinary scientific endeavour for the benefit and well-being of humanity.

20

Information Needs for Basic Research: An African Perspective

Andrew M. Kaniki
National Research Foundation, South Africa

An effective science and innovation system in any country, and globally, depends on strong basic research and higher education infrastructure. In addition to knowledge production, basic research facilities, development of human resources, and applications are critical. In the course of conducting, applying, and managing research, both researchers and managers of research and innovation have information needs. These needs must be satisfied in order for the scientists and the science innovation system to function effectively.

This paper looks at these information needs from an African perspective. However, an African perspective of the science environment is not homogeneous. Africa and its research setting constitute a complex system and African countries themselves are heterogeneous. For example, in South Africa, the science sector was fragmented due to the apartheid system of government, under which many of the institutions of higher learning were divided along racial lines, thus creating ethnic institutions of higher learning. In effect, there were different kinds of developments in those particular institutions. These issues continue to bedevil many African countries.

INFORMATION NEEDS IN BASIC RESEARCH

In discussing information needs it is useful to understand the concepts of data, information, and knowledge, and how they relate to each other. Data are the foundation of knowledge. They are a set of symbols to which rules of syntax are applied. Data are observable facts of a situation or the ingredients that make up an event. They are unstructured, isolated, and context independent, but they are capable of being integrated within a particular context. When value is added to data by either contextualizing, analyzing, or categorizing them, they are converted into information.

Information is defined as ideas, imaginative works of the mind, and data of value that are potentially useful in decision making, question answering, and problem solving. Information makes one aware of the application of available data. The acquisition of information and its appropriate application can lead a person to a state of knowledge. Being informed is central to the generation of new knowledge or the understanding of a particular situation. When information is transferred from source to recipient or from seller to buyer, it remains available to both.

Knowledge may be defined as the whole body of cognitions and skills that individuals use to solve problems. It includes theories and practical everyday rules and instructions for action. Knowledge is based on data and

information. Unlike information and data, knowledge is bound to individuals. It is constructed by individuals and represents their beliefs and causal relationships. Therefore, knowledge is dynamic; it is fluid and ever-changing. There is a lot of it that is intuitive and mutable. It is expressed through use in a moment of making, deciding, teaching, or learning. Knowledge can often be captured and structured.

In research or knowledge production data, information, and knowledge are complementary and depend on one another. The availability of relevant data that can be appropriately contextualized is critical to knowledge production. Knowledge provides a person who has the know-how, the ability, and skill to make judgments and act on given problems.

We should focus on analysis and understanding of data, instead of simply emphasizing the technology, as Chrisanthi Avgerou states.[1] If data were not contextualized, they would be meaningless. They must be transformed into information. For example, although an individual may recognize a series or set of numbers, those numbers must be organized. A group of numbers that are not organized is useless. If the set of numbers were organized in a spreadsheet, for instance, and were labeled as food production statistics over a number of years for different countries in the Southern African Development Community (SADC), the data would become useful information. The individual could then ascertain the status of food production in the southern African region. With know-how and skill the individual could manipulate those data and forecast the future of food production and usage in the region. The ability to analyze the data and put them into context is of particular importance.

Individuals, groups, organizations, and governments require information and data to make decisions. Information communication technologies (ICTs) are critical for providing different channels for communicating and sharing relevant data. Some of the information used for decision making is internalized within individuals themselves because they know the information. The individuals and data may be very localized within a particular region. Often however, human problems, organizational decisions, and government alternatives are too complex to be dealt with simply by one's internalized or localized information and knowledge. They require different types of data, information, and knowledge from a variety of sources because of the complexity and the multidisciplinary nature of the problems, particularly in developing countries. There is a need for data and information to be shared across regional borders.

An example is the issue of genetically modified maize. The decision by the southern African governments of Malawi, Zambia, and Zimbabwe to use it depends entirely on their basic understanding of what the issues are and how its use may affect their people. Without that clear understanding and contextualization it is extremely difficult to make an informed and appropriate decision. In order to provide relevant data and information researchers should understand the various needs of the people and the various contexts within which the information is used.

BASIC RESEARCH, KNOWLEDGE PRODUCTION, AND HUMAN RESOURCE DEVELOPMENT

Basic research, as defined by Frascati Manual, is experimental and theoretical work undertaken primarily to acquire knowledge of the underlying foundation of phenomena and observable facts without any particular application to use in view.[2] Research and development defined by the same manual is creative work that is undertaken in order to increase the stock of knowledge and the use of this knowledge to devise new applications.[3] It is dependent on existing knowledge, which in most instances is basic research and information.

A strong base of research within a particular country or continent, such as Africa, is critical. Basic research and the environment for research in higher education institutions provide the laboratory and the place where the expertise can be generated. People at these institutions learn the basic skills for identifying and investigating natural phenomena and scientific problems, as well as interpreting and analyzing data.

---

[1] See Chapter 10 of these *Proceedings*, "Information Technology and Data in the Context of Developing Countries," by Chrisanthi Avgerou.

[2] Organisation for Economic Co-operation and Development (OECD). 1994. *The Measurement of Scientific and Technical Activities: Standard Practice for Surveys of Research and Experimental Development—Frascati Manual 1993*, OECD, Paris.

[3] Ibid.

RESEARCH INFORMATION NEEDS IN AFRICA

As argued earlier, information needs, even among people, organizations, and countries of the same group or class, are context-dependent. There are variables, such as cognition, differences in facilities, accessibility to particular sources of information, channels that are used, and so on, that influence the nature of information needs. The degree of need for information will depend on the person's education and other factors, as well as what is available within that person's research environment.

In *Africa in the Millennium,* editors Maloka and LaRoux identify various issues that confront researchers in the political, social, and economic context of Africa.[4] The authors highlight issues of interethnic conflicts, wars, regional fragmentation of African leadership, recurrence of military governments, economic stagnation, and international marginalization. They also highlight the HIV/AIDS pandemic and the related issues of xenophobia and the migration of scientific scholars—the scientific diaspora or brain drain.

Maloka and LaRoux also highlight the Internet connectivity constraints, commodification of knowledge, and the problem of knowledge production and dissemination as issues that impact the research community in Africa. Institutions of higher learning are now being pushed to move away from basic research into the development of products that they can sell to maintain themselves. World development indicators tend to highlight these factors as a measure of the relative socioeconomic status of the African context.

A few countries, such as South Africa and Egypt, have more than one science council that participate in and fund research. In most African countries institutions of higher learning depend on government funding for their academic and research endeavors. Because of the various pressures that these institutions face and the broad roles they are expected to play (teaching, research, and community development) little or no funds are allocated for research, knowledge production, and access to the latest information in various fields. For example, many African institutions of higher learning have not renewed their journal subscriptions for many years and have no research equipment or appropriate systems for managing research. The result is that research is pushed aside because of the difficulty in obtaining funds. This creates a situation in which academics are simply teaching and in many instances indulging in consultancies rather than conducting research.

Another peculiarity of African science is the lack of science innovation policies. Most countries have not updated their science information policies, strategies, and priorities. In some countries (e.g., South Africa), however, there has been some attempt to do that.

It is against this background that African researchers and research managers operate. An effective research process requires an awareness of who is doing what, what kind of research is being conducted within a particular field, and the data and knowledge that have been generated. These data and information form the basis for further research. Similar to their counterparts in the North, African researchers and research managers should remain updated and networked with colleagues in the same field; however, in the current circumstances the majority of African researchers find data access and sharing extremely difficult to achieve.

Research and innovation managers, who operate in the same environment, find themselves in the same situation. They need information about who the experts are in particular fields and comparative data for analysis of the output of institutions in particular countries. They also need information to analyze productivity. Most important, though, research and innovation managers need applicable and adaptable research performance indicators. Nevertheless, it is almost impossible for any country in Africa to have all the data that it requires, so African researchers should participate in the international arena.

CONCLUSION

In research and knowledge production, data, information, and knowledge are complementary and interrelated. The availability of research data that can be appropriately contextualized is crucial for knowledge production. It is

---

[4]E. Maloka and E. Le Roux. eds. 2001. *Africa in the Millennium: Challenges and Prospects,* African Century Publication Series No. 3, African Institute of Southern Africa, Pretoria.

important to understand that factual data are relevant, and that African researchers and research managers should have access to them. Various databases are available, and new tools and techniques allow for data mining and pulling together the different sources of data in order to make sense of them. These databases must be interconnected and populated with relevant data from Africa itself.

The African Union and the New Partnership for Africa's Development (NEPAD) have become very important strategic initiatives. NEPAD, for instance, emphasizes the promotion of cross-border cooperation and connectivity and the utilization of knowledge currently available in centers of excellence and within each country. This assumes, of course, that the centers of excellence will continue to be maintained and remain excellent. NEPAD also emphasizes the development and adaptation of information collection and analytical capacity of the research community. It is not useful to collect data without performing sufficient analysis, correction, contextualization, and categorization. This initiative also emphasizes the generation of a critical mass of technology expertise.

At a February 2003 NEPAD workshop Dr. Ben Ngubane, the Minister of Arts, Culture, Science, and Technology in South Africa, stated that the strategic significance of the NEPAD objectives lies in their focus on strengthening regional and subregional cooperation through the use of geographical information systems, the convergence of products and standards, quality, and control, and the integration of excellence into the spheres of biotechnology and natural sciences. It is critical that within the African region we contribute to the data generated around the world.

21

International Transfer of Information in the Physical Sciences

R. Stephen Berry
University of Chicago, United States

INTRODUCTION

Open access and the ready exchange of data and other information are at the heart of the normal processes of science. This is perhaps even truer in the physical sciences than in other disciplines because they may offer fewer direct pathways to profitable applications. Obvious symptoms revealing this tradition are the longstanding practice of circulating preprints and other early-stage information among colleagues and the practice of presenting results in symposia and other conferences well before the work has been submitted for publication. Electronic media have made such access steadily easier and cheaper, and consequently have made it possible for science to progress faster and at a higher level of scholarship.[1] We are in a period of adaptation and learning; this discussion starts with the perspective that we are trying to find effective ways to maintain procedures and values we know and trust while making optimal use of the powerful tools of electronic media and open access.

The following discussion will describe the kinds of information that physical scientists share, how they go about sharing that information, how the modes of sharing are changing, what the larger context is in which this sharing takes place, and then to the central point of this discussion, what challenges and problems face the scientific community and the infrastructure that supports it. Finally, I will introduce a proposal for one approach that may be a useful way of adapting to the evolving world of scientific communication.

WHAT PHYSICAL SCIENTISTS COMMUNICATE AND HOW THEY DO IT

The most obvious material that all scientists communicate is the substance contained in their formal publications. These papers in traditional journals go through a screening in the form of anonymous peer review that has set the standard threshold for acceptable distribution. There is a sort of tacit certification that goes with publication in these journals. However it must be understood that the threshold for acceptance is a relatively low (although sometimes capricious) one. The normal journal review process is certainly not capable of uncovering deliberate fraud, and very rarely can reveal subtle errors or inconsistencies that would require extensive work to discover.

[1]The author is not aware of any documentation to show the extent to which scientific progress has accelerated or current publications refer more extensively to relevant prior work than before electronic communication became widespread. The term "have made it possible" is used quite deliberately.

Acceptance for publication in a journal implies that the work appears to be consistent and plausible enough to merit the attention of the relevant community of scientists and a topic for further discussion and investigation; it is neither clearly wrong nor outlandishly foolish, and may well be right. Reviewers, realistically, can devote only very limited effort to each manuscript they receive, so they screen the papers using a relatively low standard for acceptance: no obvious errors and relevance and interest for the audience of the journal.

Scientists also communicate data, sometimes just as single facts or numbers, sometimes as extensive sets of tabulated results from experiment or theory. At one extreme are the new values of some freshly measured quantity; at the other are the carefully constructed tables of critically evaluated data such as those published in the *Journal of Physical and Chemical Reference Data*. An illustration of this latter category is the data sets maintained by the National Institute of Standards and Technology, which are most easily obtained on its Web page.[2] There one can find the best current values of many quantities, for example those of the fundamental standards such as the speed of light, the masses of elementary particles, and Avogadro's number. But there are also vast sets of data available in other open databases. One that overlaps the physical and biological sciences is the Protein Data Bank,[3] a rich and fast-growing source of information about the sequences and structures that have been established and to some extent the level of uncertainty of the data. We should not lose sight, however, of the importance of one scientist relating to another the frequency of some critical, newly measured spectral line, or the strength of the bond between a protein molecule and an inhibitor molecule that sticks to the protein.

In this context it is useful to distinguish simple data repositories from data in critically evaluated databases. The latter in the United States are considered worthy of copyright; hence they have value the law considers worth protecting. Unevaluated data deposited as they are generated may be useful but dangerous to use indiscriminately. For example, most scientists familiar with the sequences in the human genome database believe that there are many errors in those data and that considerable caution should be exercised by anyone trying to use those data. One might go so far as to say that unevaluated data deposited in an indiscriminate repository should be given no legal protection at all, until they have been scrutinized critically.

Traditional journal publication and the almost-as-traditional preprint circulation have been the most important archival modes of communication. Apart from these, presentations at conferences provide another important mode of communicating scientific information. These presentations range from talks at very large meetings of professional societies to smaller, more focused meetings, such as Gordon Conferences, to the very small working groups that have met, for example, at the Aspen Center for Physics and the Telluride Summer Research Center. While these have been very important vehicles of communication, they rarely have had archival functions. In fact, Gordon Conference rules prohibit any record of the discussions or even the formal presentations; even quoting a presentation requires the permission of the person being quoted. Conference proceedings have in many instances been published, but unless the vehicle for publication is a standard journal and the issues containing the conference proceedings are circulated as normal issues, the proceedings become almost invisible to scientists later. There has certainly been disillusionment among researchers about publishing work in conference proceedings. On the other hand, the information exchanged among participants, especially in the smaller meetings, plays a very seminal role in moving science ahead.

The most obvious changes of the past 20 years have been the increase in electronic modes of archiving, communicating, and accessing scientific information. There has probably been an increase in the frequency of small, specialized conferences and workshops as well. Electronic communication has its most known manifestations in electronic versions of journals and in less conventional electronic forms of communicating— publishing— completed work. Martin Blume distinguishes "Publishing" from "publishing"; the former implies appearance in a conventional journal, with reviewing and editing included in the process of "Publishing," while "publishing" includes both that mode and also posting on an electronic server such as the arXiv[4] that imposes no reviewing or editing and simply accepts what a scientist submits.

[2]See http://www.nist.gov.
[3]See http://www.rcsb.org/pdb/.
[4]See http://arXiv.org, formerly xxx.lanl.gov.

There is another very important but more subtle change that the electronic communication of scientific information has engendered, particularly regarding international activities. The fast, relatively reliable, and marvelously convenient, boundary-free medium of e-mail has almost erased national and other geographic boundaries between interacting scientists. The extent of international collaboration has expanded enormously since the late 1980s because such communication is available. It is equally easy for a scientist in Chicago to collaborate with a colleague in Moscow or in Madison. The exchange of information, whether it be informal ideas, elaborate animations, funding proposals, manuscripts in preparation or completed, even drafts of full books, becomes a natural day-to-day process. We have already reached a point at which we take such communication for granted. If an institution's Internet server crashes for a day, its research program comes to a panicked halt. A researcher may communicate with colleagues, many of them collaborators, in five or six countries in a single typical day. An especially relevant aspect of this internationalization is that many of the scientists participating in these exchanges are in developing countries, where the researchers now typically have access to computers, and hence to electronic archives but not to extensive traditional libraries. The importance of this point will become explicit when we examine how journals are distributing their publications.

THE LARGER CONTEXT

The scientific research we have been describing is largely conducted in academic settings, federal laboratories, and research institutions devoted to basic science. This research is supported largely by funding from federal sources and not-for-profit foundations. In the biological sciences the proportion from private sources, particularly pharmaceutical firms and others in related areas, is somewhat higher than in the physical sciences.

The justification for this support is the collection of public goods produced by the research, goods that would not come into existence if it were not for federal and foundation funding. This aspect of basic research has been emphasized previously, but it should be stated again and again to reach wider audiences.[5] A public good is a good whose value does not diminish with use. Public goods produced from science typically increase in value with use; this is one of two crucial characteristics of traditional science that must be a dominant factor in how we select any policy for science. Science's product as a public good implies that any institution that funds research for the purpose of producing public goods carries the responsibility for seeing that the results of the research are disseminated. Dissemination is necessary for the public goods to emerge from research. Because the public goods from research amplify with increasing use, the benefits of dissemination typically produce a marginal return that may increase rapidly beyond the initial investment in that dissemination.

The mechanisms of dissemination discussed in the previous section are undergoing dramatic changes. Fifty years ago, the symbiotic relation between journal publishers and scientific researchers provided a healthy way for science to disseminate and archive its products. The publishers—adding value to the material by editing, distributing, indexing, and archiving it—were able to profit monetarily while the scientists profited through access to the information. The number of scientific journals published in the United States increased steadily and exponentially from about 1840 until the late 1990s, when the growth rate slowed a bit.[6] During the rapid expansion of basic scientific research during the 1960s and 1970s, many new, specialized journals appeared. Consequently, university libraries, federal laboratories, and industrial research centers subscribed to many more journals than they had prior to 1960. When budgets could no longer keep pace with rising numbers of journals and rising subscription prices, libraries became more selective and began to drop subscriptions and form coalitions to share the more specialized and less frequently used journals.

The situation changed again with the evolution of electronic distribution. Electronic storage of data, especially large-volume databases, was being used by such agencies as the National Library of Medicine and National

[5]See National Research Council. 1997. *Bits of Power: Issues in Global Access to Scientific Data*, National Academy Press, Washington, D.C.; K. Dam. 1999. *The Changing Character, Use and Protection of Intellectual Property,* Stiftung, Deutsch-Amerikanishes Akademisches Konzil (DAAK) Symposium Band 11, Bonn, Germany, pp. 17-36; and R. S. Berry. 2000. *"Full and Open Access" to Scientific Information: An Academic's View*, Learned Publishing, Baltimore, pp. 37-42.

[6]Carol Tenopir and Donald W. King. 2000. *Towards Electronic Journals*, Special Libraries Publications, Washington, D.C.

Aeronautics and Space Administration in the early 1980s. Electronic access to e-mail became widespread in the 1980s. Journals first began to use electronic methods in doing their composition; the next step was introducing modes of accessing individual papers.[7] Journals then moved toward greater and greater use of electronic tools, starting with search procedures based on keywords; next, posting images of abstracts and then full manuscripts; putting up full texts in searchable form; and then to making available supplementary material that did not appear in the paper version of articles. The next stage, something envisioned in the mid-twentieth century but only realized in the late 1990s, was the posting of full archives of all issues of journals. Naturally, this question arose: what credentials would give a user access to the information put into such computer-accessible files? This issue is one I will explore further. At this point let us only point out that publishers have sought to give access to as wide an audience as each of them considers safe within the limits of stability of their publications. At issue are what those limits are, what those limits should be, and how those limits are affected by the procedures under which a publication operates; for example, some journal publishers have made their publications available without charge in developing countries.

CHALLENGES AND PROBLEMS

With the development of electronic distribution and the move toward open access, journal publishers—both private firms and professional societies—faced a major challenge to their financial stability. Professional societies adopted policies ranging from the enthusiastic acceptance of open access (e.g., the American Physical Society) to rather strongly protective positions (e.g., the American Chemical Society). This should be considered a healthy way the larger society can conduct experiments, and let the experiments play out. To go one step further with this logic: at this time in the evolution of new practices to adapt to new technology and its consequences, it is particularly important that any legislation and regulation that we adopt be permissive rather than restrictive. Restrictive laws and regulations at such a stage are extremely counterproductive, inhibiting competition among alternative courses of action. Any action such as the European Commission's Database Directive that makes some courses difficult can only be interpreted as retrograde protection of narrow interests at the expense of the benefit of the larger society. If the more protective course were to win out over more open modes, so be it, but to prevent the possibility of one mode competing against others is simply dangerously bad economics. One possible scenario that might become reality splits the community of scientists and their publishers into one group that supports and uses open distribution of information and another group that follows a policy of strong protection; in such a case it would be very likely that the former group would simply exclude the latter from its regular attention, thereby neglecting the works produced by the more protectionist faction.

There are two challenges most apparent for any new mode of operation within the electronic environment. The first is deciding how material intended for publication should pass certification, that is, what do we do about reviewing? The second is determining how we support the components of the scientific enterprise that have been paid for previously through the various forms of journal support. This funding includes the costs of publication, distribution, and archiving and in some cases the primary support of professional societies.

Peer review has been the subject of a recent discussion by Paul Ginsparg.[8] Here the range of solutions is likely to remain broad because of the range of practices among the sciences and the differences in levels of concern regarding the impact of published material on the readers. The practice of circulating preprints, a long-standing mode in high-energy physics, has given that community a sense of confidence in the effectiveness of postpublication review. Open commentary serving as postpublication review is a widely practiced way for the communities using the online arXiv to provide such "reviews" of articles published there. Ginsparg enjoys telling one example of an article posted on the arXiv at the same time it appeared in *Physical Review Letters* as a refereed paper. Within a day several contributors to the arXiv showed in comments published there and appearing with the

[7] An extensive history of electronic methods in journals, and the prescient earlier publications on the subject, are provided in Tenopir and King, 2000.

[8] See http://arXiv.org/blurb/pg02pr.html.

original article that there were fundamental flaws in the published work. In other words, the online commentary provided a much more stringent, highly relevant review than did the normal anonymous refereeing through the journal. On the other hand, in the biological sciences, researchers worry that nonscientists would read unreviewed articles posted on an open source and try to use health care approaches that they find there. Without any evaluation of soundness many biomedical scientists feel this would be dangerous.

Ginsparg proposes a procedure of systematic review for articles deemed worthy of particular scrutiny sometime after they have been put up on an electronic archive. His procedure would displace formal reviewing to a review stage after the normal audience had seen publications but before they went through any formal review. One motivation for such a procedure would be to reduce the costs of publication; reviewing all submitted papers as it is now done is one of the more costly stages of publication. The two-stage process would have the advantage of giving the professional audience rapid access, through the arXiv or a parallel counterpart, and would still provide certification for the relatively small fraction of papers that then go through formal review.

The other major problem facing scientists and publishers is the issue of payment to present results of research to its audience. The primary source of financial support of printed journals is subscriptions bought by libraries. The prices for individual subscriptions of many journals are essentially the marginal costs of printing and mailing the journals. Commercial journals are priced to generate profits at a level chosen by the publisher. Journals published by professional societies (rather than by commercial publishers acting as agents or contractors for the societies) are priced considerably lower than those published by commercial houses. Even so, many professional societies use income from journal subscriptions as important sources of revenue for the activities of the societies. Such organizations are quite naturally very protective of the income they receive from publications.

The world of science publishing is divided on how to treat the posting of articles on electronic servers, particularly on servers available without charge to readers. Commercial publishers that post the contents of their journals online typically exact a charge from subscribers, by way of institutional subscriptions that allow anyone from a particular institution to download articles or individual subscriptions that allow a particular user to have access. Some professional societies also follow that policy. Other journal publishers, such as the American Physical Society and the U.S. National Academy of Sciences, make the contents of their papers available without charge after some delay. The electronic arXiv makes papers available immediately.

Journal policies are sharply divided regarding whether posting content on an open-source site, such as the arXiv, or presenting the work at a conference should influence the acceptability of a paper for publication. Some journals, such as those of the American Physical Society, actually cooperate with open-source providers. Others allow open-source distribution only after a specified interval following publication. Still others consider presentation of work at a conference prior to submission of a manuscript as grounds for disqualification of the manuscript for publication. The publishers that are reluctant to post published works on a no-fee basis are fearful that they will lose many subscriptions by such posting.

Whether this is a correct assumption will presumably be demonstrated by the other publishers who do allow no-charge access. This is an important experiment, whose consequences should be examined and evaluated. It will not be sufficient to look only at the net revenue changes of the firms that post their articles. It will also be important to observe how the journals in one category compete with their parallels in another. Wherever there are competing journals in the same field offering authors the option of choosing one kind of publisher or the other, there is an opportunity to learn whether the posting policy of the publisher influences the choice of the author about which vehicle to use for publication. It will also be important to evaluate the impact of the two categories of journals; will no-fee electronic posting have a significant effect on the extent to which readers use a journal? We can easily predict with confidence that journals that distribute their publications without charge to developing countries will have significant impact in those places, but what about such distributions in developed countries?

One facet of the financing issue is the question of which procedure should be used to pay the publisher. Subscriptions, one form of user fees, are the most widely used method now. During the years when electronic publishing was still a far-off goal, the readers' pay-per-article model was discussed frequently. An altogether different approach is the author-based payment. Page charges were a form of this mode but were never constructed to cover more than a relatively small fraction of the total cost. Page charges were so unacceptable to the author community that they have become voluntary for many journals. If an author has to decide between paying a

voluntary page charge and buying a new computer, the decision is easy and obvious. The funds in a grant that are at the investigator's disposal are simply too valuable to be used to publish a paper, if that paper can be published without charge. Yet, as previously mentioned, the funder of that grant has a responsibility to see that the results of the research supported by the grant are distributed and used. This paradox faces the community of scientists, publishers, and funders of research right now.

A MODEST PROPOSAL

This discussion concludes with a proposal to resolve the problem of paying for publication of scientific papers. The cost of publishing, high as it seems in the context of open access to scientific data and information, is a small fraction of the total cost of the research the publications describe. If a traditional journal can maintain high quality with low subscription costs and without its publisher charging authors so that the journal remains healthy and competitive, there is no need for any change. However the journal publishers, particularly the professional societies whose primary motivation is to serve their scientific memberships, see financing as a goal secondary to keeping the journals and the societies healthy. Their not-for-profit roles determine priorities that are quite different from those of commercial publishers. Consequently, a professional society is likely to regard electronic posting differently than a commercial publisher, and is likely to have a different kind of motivation toward actions that might threaten traditional revenue sources.

The method we propose for paying for the publication of scientific papers is a fallback that will ensure that the funder's goals are achieved from the perspective of the scientific community, which is the heaviest user of the published results. We propose an author-based fee system, but one in which the author is not expected to cover the publication costs using a research grant. Instead, we propose that the journal publishing an article bill the funding agency or foundation directly for the costs of editing, reviewing (so long as the paper goes through review), indexing, distributing, and archiving that are beyond what subscription revenue can cover. Direct billing would ensure that journals would survive however deeply they commit themselves to electronic accessibility and however much that accessibility cuts into subscriptions. This procedure was used, in fact, for some years by one U.S. foundation for publications generated in the research projects it supported.

This method might appear superficially as a direct subsidy, but it is a rechanneling of funds to reduce transaction costs and achieve the goals of the researchers and their financial supporters. At present many of the funds that cover indirect costs of a grant are used to support institutional libraries. If subscription costs could be held down by shifting publication costs significantly to an author-based charging system, then some overhead funds could be included in a direct publication charge and would apply specifically to the user community of those publications.

It would be reasonable to have some upper limit on what could be charged. The amount could be based on figures developed by professional societies working with funding agencies. If a publisher were to charge more than that minimum, the author would be expected to pay that difference with discretionary funds, presumably from a grant. This might happen, for example, with very-high-prestige journals of commercial publishers.

It would be quite possible for some journals to opt not to participate in such a scheme. In fact, competition with other alternatives would be a desirable way for the method to be tested. It would require collaboration of the funding agencies, but it might be possible to adjust the distribution of funding so that the total budget for support of research could be nearly constant while the distribution of scientific results could become much more effective.

22

Access to Scientific Information: The Ukrainian Research and Academic Network

Mikhail Zgurovsky
National Technical University of Ukraine

The Ukraine and Russia have a long and strong history in the areas of cybernetics, mathematics, and computer sciences. In 1952 the Ukraine created the third computer in the world, after the United States and Great Britain. Yet, all of the social and economic problems of the last decade have greatly reduced these investigations, especially in technology.

THE UKRAINIAN RESEARCH AND ACADEMIC NETWORK

In the course of the last six years Ukrainian science and education have been actively developed, and there is a Ukrainian scientific and educational information portal. The portal is a powerful information gateway. The main segment of this portal is the scientific and educational computer network: the Ukrainian Research and Academic Network (URAN). URAN was created in two stages. The backbone was implemented from 1996 to 2000 with the basic nodes in six Ukrainian cities. In 2002 the second stage was implemented and the regional nodes were created in six more cities.

The topology of URAN is star-shaped with reserve segments that are determined by application of asynchronous-transfer-mode networking technology. URAN uses various types of channels, including the main ground-based channels of the Ukrainian telephone company and its own satellite channels. The URAN has a three-level architecture. The first level includes the central node in Kiev, the main optical fiber and satellite channels, and the data transfer connected to the Internet. The second level includes the network regional nodes. The third level comprises information infrastructure and incorporates networks of universities, academic institutions, and scientific libraries. This is where users access the URAN resources. The significant progress in integration of the Ukrainian scientific and educational information portal with European and global information was made by its connection to the international educational and academic network, GEANT, a detector description and simulation tool. The main optic fiber channels of this network have a data transfer rate of 10 gigabits per second. In November 2002 Ukraine was the first of the Commonwealth of Independent States countries to be connected to GEANT via Ukrainian telecom channels at a rate of 34 megabits per second.

The second segment of the Ukrainian information portal includes centers with educational and informational software that provide access to students. This distance learning system coordinates and provides access to organizations, distance learning and professional orientation centers, educational and scientific institutions, developers,

and students of the system using the resources of URAN. The methodology of distance learning is based upon the principles of shells. In such shells systems, which have their own internal structure and connections, only the content part of the shells changes, namely, information. Despite a variation in information other general system models of the shell do not change. At present there are 15 regional nodes of distance learning in the Ukraine, which are connected in the integrated medium by means of the URAN system.

The third segment of the Ukrainian information portal provides access to the electronic information resources of science and education in accordance with the three levels of the computer network—Internet, intranet, and corporate networks. This segment is connected to regional laboratories with remote access to information resources in the fields of economics and management, ecology, medicine, biology, physics, mathematics, modeling of complex processes, telemedicine, and other fields. These resources include online electronic libraries from various organizations, universities, academic institutions, the Ukrainian National Academy of Sciences, public libraries, and distance learning centers. The network of online electronic libraries includes library resources in scientific and educational fields. The system is built upon unified operational software, hardware, and technical platforms.

The fourth segment of the informational portal is a national educational information system called OSVITA. This system solves the educational management problems, such as the problems of automation of information collection and processing and the collection of state documents on education. OSVITA is a complex grouping of administrative legal, software, and hardware facilities. The system makes it possible to create a unified information infrastructure to process data on education, preserving their validity and integrity, and providing reliable information protection mechanisms. The OSVITA system ensures that complex information production systems used to create state documents on education will function.

URAN APPLICATIONS

An application of the Ukrainian information portal involves solving the problem of ecological monitoring of the Chernobyl nuclear power plant, as well as the development of the telemedical channel for the diagnosis and treatment of those working at this plant. One of the problems is the necessity for continuous control over the area above the plant. There are a number of regional centers monitoring ecological conditions in the area. The continuous monitoring of ecological conditions in the Ukraine shows that the plant zone itself and considerable adjacent territories are still polluted with radioactive substances such as plutonium isotopes. Such complex industrial factors have no analogies from the point of view of their harmful impact upon human health.

The power engineering specialists working in the Chernobyl nuclear power plant live approximately 30 km from Chernobyl. There are 4,363 people working at the reactor. Of those, 1,375 are women. Practically all the inhabitants of the city are in contact with ionizing radiation conditions. More than 84 percent of the inhabitants suffer from the Chernobyl disaster. These people need special, high-quality medical service and exhaustive consulting assistance. The obstacle is insufficient technical equipment in patient care institutions and remoteness from the regional center.

The purpose of the monitoring system is also to analyze the state of health of various professional groups while taking into account the risk factors. The system will make it possible to work out individual programs of health recovery. The absence of such a system is one of the reasons that approximately 600,000 people have become invalids from the Chernobyl catastrophe. A number of measures aimed at protecting the plant's personnel against harmful impacts and risk factors will be implemented. Medical monitoring is necessary to establish the expert diagnostic system and the system of remote transfer of the analysis and research data. It will make it possible to carry out consultations by means of telecommunication.

The Chernobyl telemedicine system consists of two centers. One is the diagnostic and rehabilitation center, which is situated in Slavutych City. It monitors the main medical factors of employees of the Chernobyl power plant and members of their families, in accordance with individual programs. These data are transferred by the URAN system to the international telemedicine center, which is located in Kiev, at the National Technical University of Ukraine—Kiev Polytechnic Institute. Highly qualified specialists in the field of medicine work in this center. They analyze the data obtained on each patient and work out recommendations for prevention and

treatment. The project is intended for three years and its estimated cost is about U.S. $300,000. This project is financed from the municipal budget of Chernobyl.

Development of the Ukrainian information portal will help solve the problems of access to the digital information resources of different sectors of science and education in the Ukraine.

SESSION 5: INNOVATIVE MODELS FOR PUBLIC-DOMAIN PRODUCTION OF AND OPEN ACCESS TO SCIENTIFIC AND TECHNICAL DATA AND INFORMATION

23

Introductory Remarks by Session Chair

Dominique Foray
Centre National de la Recherche Scientifique, France

Many people agree that public-domain production of and open access to scientific and technical data and information are key features in the process of scientific discovery, invention, and innovation. They constitute a fundamental institutional arrangement and requirement for economic growth and development. The shared collection of data and information provides the building blocks for innovation and invention, and those free journeys through information space are a key factor for the scientific enterprise.

It would be a mistake, however, to think that the old institutional framework that was mainly based on government support must be maintained in the hope that it will continue to support the production of public-domain information. While the old framework did well from the postwar period until the 1980s, the great challenges and issues that Session 5 is addressing concern the new types of institutions and mechanisms that will create favorable conditions for the provision of this kind of public good in the area of basic research in the knowledge economy. The important phrase in the title of this session is "innovative models." The speakers will try to be innovative and think about the new kinds of institutions that are required, given that the old framework is no longer sustainable.

The first step in the quest for innovative models is to recognize that the public domain, which means public spaces that are conducive to efficient knowledge dissemination, does not necessarily mean "government control" or "public sector." In particular, collective actions are important in supporting the public domain and open access. Some social scientists call this category of collective action inherently public property, meaning public property that is controlled neither by the government nor by private actions. It is probably this category of public property that will constitute the framework for the revival of the public domain of the knowledge-based economy.

Of course, a variety of innovative models are possible and should be discussed. They include mechanisms aimed at maintaining a public space and open knowledge commons by voluntary means; promoting an ethos of information sharing; open-source movements, supported by very complex incentive structures and by a legal framework that protects the public domain; and by the transfer of databases, information, and knowledge from the private to the public sectors. It is also important to think about mechanisms that use the legal framework of the intellectual-property-right system to impose access rights, such as compulsory licensing and price discrimination.

24

A Contractually Reconstructed Research Commons for Scientific Data: International Considerations[1]

Jerome Reichman and Paul Uhlir
Duke University Law School, United States, and The National Academies, United States

The importance of public-domain data for scientific research is taken so much for granted that it is difficult to identify the precise boundaries of this domain, to describe its operations, and to evaluate the normative and legal infrastructure that supports it. Individual researchers know a little about the public domain as it affects their discipline, but very few understand how the overall system works.

Two new challenges emerge in this context. The first is the advent of digital networks, which transforms the traditional modes of exchanging scientific data and increases the payoffs that both science and industry can draw from open access. The other challenge is a growing array of economic, legal, and technical pressures that threaten to impede and even to disrupt the continued operations of this same public domain for scientific data.[2]

Our investigation reveals that the policy of open access to public research data rests on a surprisingly fragile foundation in both legal and normative senses. As universities and scientists increasingly aim to commercialize their research products and profit from them, their willingness to exchange data, along with other research tools, has been seriously compromised.

There is evidence that informal exchanges of data in some fields, such as biomedical research, between individual scientists and laboratories, usually at the prepublication stage, have become severely compromised, with as much as 50 percent denials of requests for information.[3] At the same time, interuniversity exchanges of data are subject increasingly to high transaction costs, delays, and a growing risk of anticommons effects,[4] that is, too many intellectual property rights and commercial interests making it difficult to put together complex transactions. As relations between universities and industry become more intense there is a growing ability of the industrial partner to impose the terms of exchange on the universities.

---

[1]This presentation is based on a recently published article. See J. H. Reichman and Paul F. Uhlir. 2003. "A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment," 66 *Law and Contemporary Problems*, Winter-Spring, 315-462.

[2]For a detailed description of the pressures that may impact the public domain for scientific data, see ibid., pages 361-416.

[3]Eric G. Campbell et al. 2002. "Data Withholding in Academic Genetics: Evidence from a National Survey," 287 *JAMA* 473-80. Moreover, 10 percent of all postpublication requests for additional information were reportedly denied.

[4]See, e.g., Stephen M. Maurer. 2001. "Inside the Anticommons: Academic Scientists' Struggle to Commercialize Human Mutations Data, 1999-2001," paper given at the Franco-American Conference on the Economics, Law, and History of Intellectual Property Rights, Haas School of Business, University of California at Berkeley, October 5-6.

The advent of strong new intellectual property rights, such as the European Commission's directive on the legal protection of databases, strain this very delicate situation and can disproportionately affect scientists' access to and use of factual data historically in the public domain. This database right allows the scientist or the university to publish the article and still retain ownership of the data after publication. Similarly, this database right makes it possible for scientists or universities to apply for patents and disclose the data in the patent but still retain ownership of the supporting data even after the patent expires.

Such a database right when combined with other economic and legal pressures could become the hub of a growing enclosure movement that could progressively fence off the public domain for scientific data. It could reduce the flow of data as an input into national systems of innovation, and into the international system. Policy makers therefore must take steps now to address these challenges and to ward off this threat of undue enclosure in order to be able to exploit the new benefits of digitally linked databases.

THE CHALLENGE FOR SCIENCE

These trends could elicit one of two types of responses. One response is essentially reactive, in which the scientific community adjusts to the pressures as best it can without organizing a response to the increasing encroachment of a commercial ethos upon its upstream data resources. The other would require science policy to confront the challenge by formulating a strategy that would enable the scientific community to take charge of its basic data supply and to manage the resulting research commons in ways that preserve its public-good functions without impeding socially beneficial commercial opportunities.

Under the first alternative the research community can join the enclosure movement and profit from it. Thus both universities and independent laboratories or investigators that already transfer publicly funded technology to the private sector can also profit from the licensing of databases. In that case data flows supporting public science will have to be constructed deal by deal with all the transaction costs this entails and with the further risk of bargaining to impasse. The ability of researchers to access and aggregate the information they need to produce discoveries and innovations may be compromised both by the shrinking dimensions of the public domain and by the demise of the sharing ethos in the nonprofit research community, as these same universities and research centers increasingly see each other as competitors rather than partners in a common venture. Carried to an extreme this competition of research entities against one another, conducted by their respective legal offices, could obstruct and disrupt the scientific data commons.

To avoid these outcomes the other option is for the scientific community to take its own data management problems in hand. The idea is to reinforce and recreate, by voluntary means, a public space in which the traditional sharing ethos can be preserved and insulated from the commoditizing trends identified above. In approaching this option the community's assets are the formal structures that surround government-funded data and the ability of the funding agencies to regulate the terms on which data are disseminated and used. The first programmatic response would look to the strengthening of existing institutional, cultural, and contractual mechanisms that already support the research commons, with a view to better addressing the new threats to the public domain identified above. The second logical response is to react collectively to new information laws and related economic and technical pressures by negotiating contractual agreements between stakeholders to preserve and enhance the research commons.

In the United States the government generates a vast public domain for its own data by a creative use of three instruments: intellectual property rights, contracts, and new technologies of communication and delivery. By long tradition the federal government has used these instruments differently from the rest of the world. It waives its property rights in government-generated information, it contractually mandates that such information should be provided at the marginal cost of dissemination, and it has been a major proponent and user of the Internet to make its information as widely available as possible. In other words, it has deliberately made use of existing intellectual property rights, contracts, and technologies to construct a research commons for the flow of scientific data as a public good. The unique combination of these instruments is a key aspect of the success of our national research enterprise.

Now that the research commons has come under attack in the United States and elsewhere, the challenge is not only to strengthen a demonstrably successful system at the governmental level but is also to extend and adapt this methodology to the changing university environment and to the new digitally networked research environment. In other words, universities, not-for-profit research institutes, and academic investigators, all of whom depend on the sharing of data, should stipulate in their own treaties or contractual arrangements to ensure unimpeded access to and unrestricted use of commonly needed raw materials in a public or quasi-public space, even though many such institutions or actors may separately engage in transfers of information for economic gain. This initiative in turn will require the federal government as the primary funder—acting through the science agencies—to join with the universities and scientific organizations to develop suitable contractual templates that can be used to regulate or influence the research commons.

Implementing these proposals will require nuanced solutions tailor-made to the needs of government, academia, and industry in general and to the specific exigencies of different scientific disciplines. The following sections briefly summarize my proposals for preserving and promoting the public-domain status of government-generated scientific data and of government-funded scientific data, respectively. In this scenario data flows between scientists will have to be constructed deal by deal, with all the risks of bargaining to impasse that we already see in biotechnology today.

PROPOSALS FOR THE GOVERNMENT SECTOR

To preserve and maintain the traditional public-domain functions of government-generated data the United States will have to adjust its existing policies and practices to take account of new information regimes and the growing pressures for privatization. At the same time government agencies will have to find ways of coping with bilateral data exchanges with other countries that exercise intellectual property rights in their own data collections.

We do not mean to imply a need to totally reinvent or reorganize the existing universe in which scientific data are disseminated and exchanged. The opposite is true. A vast system of institutional mechanisms for the diffusion of scientific data as a public good—especially government-generated data—exists and continues to operate, and much government-funded data emerging from the academic communities continues to be disseminated through these well-established mechanisms.

In the European Union a strong database right exists and governments already exercise this right. Wise statesmanship would require these governments to renounce part of the right to better promote scientific progress. Governments in all countries, however, should consider imposing contractual templates in their own relations with the private sector. When they license data to the private sector for exploitation, they should ensure that the private sector will not harm scientific activities when it exploits the data. Contractual templates are needed to govern those kinds of relations and to help ensure the more efficient operations of the public-sector research community. These underlying contractual templates could implement the following research-friendly guidelines:

1. A general prohibition against legally or technically hindering access to the data in question for nonprofit scientific research and educational purposes;
2. A further prohibition against hindering or restricting the reuse of data lawfully obtained in the furtherance of nonprofit scientific research activities; and
3. An obligation to make data available for nonprofit research and educational purposes on fair and reasonable terms and conditions, subject to impartial review and arbitration of the rates and terms actually applied.

Governments from high-protectionist areas and from low-protectionist areas will have to come to some sort of international treaty that will allow minimum protection of databases without requiring every country to adopt the highest form of protection while keeping in mind the special requirements for the circulation between countries with different levels of scientific data.[5]

---

[5]See Professor Reichman's review of database protection in the international context, which proposes this type of treaty. J.H. Reichman. 2002. "Database Protection in a Global Economy," *2002 Revue Internationale de Droit Economique*, pages 455-504.

PROPOSALS FOR THE ACADEMIC SECTOR

The primary focus of this presentation is on proposals for the academic sector, because so much of the data are government funded and go through the universities. Therefore, they already benefit from a rudimentary regulatory regime. We suggest that science policy should treat data produced with government funds as a collective resource for research purposes and offer proposals that deal with how to do that.[6]

It is possible to differentiate between two "zones" of government-funded data. The first is a zone of formally regulated data exchanges, for which the regulations are imposed by the funding agency and generally kick in at the time of publication. The second is a zone of informal data exchanges, which typically run in the prepublication research phase, as well as in situations in which the terms of making data available are not formally specified in a research contract or grant. In Europe it would also be in the postpublication phase because of the existing database right. The ability of government funding agencies to influence data exchange practices will be much greater in the formal zone rather than in the informal one.

The Zone of Formal Data Exchanges

When no significant proprietary interests come into play, the optimal solution for government-generated data and for data produced by government-funded research is a formally structured archival data center also supported by government. Many such data centers have already been formed around large facility research projects. Building on the opportunities afforded by digital networks, it has now become possible to extend this time-tested model to highly distributed research operations conducted by groups of academics in different countries.

The traditional model entails a "bricks-and-mortar" centralized facility into which researchers deposit their data unconditionally. In addition to academics, contributors may include government and even private-sector scientists, but in most cases the true public-domain status of any data deposited is usually maintained. Examples in the United States include the National Center for Biotechnology Information, directly operated by the National Institutes of Health, and the National Center for Atmospheric Research, operated by a university consortium and funded primarily by the National Science Foundation. Hundreds of such data centers already exist.

A second, more recent model enabled by improved Internet capabilities also envisions a centralized administrative entity, but this entity governs a network of highly distributed smaller data repositories, sometimes referred to as "nodes." Together the nodes constitute a virtual archive whose relatively small central office oversees agreed technical, operational, and legal standards to which all member nodes adhere. Examples of such decentralized networks, which operate on a public-domain basis in the United States, are the National Aeronautics and Space Administration's Distributed Active Archive Centers under the Earth Observing System Program and the Long Term Ecological Research Network funded by the National Science Foundation.

These virtual archives, known as "federated" data management systems, extend the benefits and practices of a centralized bricks-and-mortar repository to the outlying districts and suburbs of the scientific enterprise. They help to reconcile practice with theory in the sense that the investigators—most of whom are funded by government anyway—are encouraged to deposit their data in such networked facilities. The very existence of these formally constituted networks thus helps to ensure that the resulting data are effectively made available to the scientific community as a whole, which means that the social benefits of public funding are more perfectly captured and the sharing ethos is more fully implemented.

At the same time some of the existing "networks of nodes" have already adopted the practice of providing conditional availability of their data, a feature of considerable importance for our proposals. "Conditional availability" means that the members of the network have agreed to make their data available for public science purposes on mutually acceptable terms, but the members also permit suppliers to restrict uses of their data for other purposes, typically with a view to preserving their commercial opportunities.

The networked systems thus provide prospective suppliers with a mix of options to accommodate deposits ranging from true public-domain status to fully proprietary data that have been made available subject to rules the

[6]For detailed information on these proposals, see Reichman and Uhlir, op. cit., note 1, pages 425-456.

member nodes have adopted. The element of flexibility that conditional deposits afford make these federated data management systems particularly responsive to the realities of current university research in areas of scientific investigation where commercial opportunities are especially prominent.

Basic Proposals

Several proposals are suggested for universities, and include the following:

- Develop interinstitutional agreements and cooperative institutional approaches to ensure unimpeded access to and liberal uses of scientific data and information in a not-for-profit research commons, while allowing for commercial exploitation of those resources in the private sector, when this is considered necessary and appropriate.
- Develop model contractual provisions for interuniversity and interresearcher relationships and for cooperative research with the private sector that protect access to and unrestricted use of publicly funded research data by not-for-profit scientists.
- Vigorously promote nonexclusive licensing by authors of their scientific articles to scientific, technical, and medical journals rather than assigning exclusive copyrights.
- Initiate and review pilot projects for certain disciplines or categories of data to test the results.

Ancillary Considerations

We are aware that considerable thought has recently been given to the construction of voluntary social structures to support the production of large, complex information projects. Particularly relevant in this regard are the open-source software movement that has collectively developed and managed the GNU/Linux Operating System and the Creative Commons, which seeks to encourage authors and artists to dedicate some or all of their exclusive rights to the public domain. In both these pioneering movements, agreed contractual templates have been experimentally developed to reverse or constrain the exclusionary effects of strong intellectual property rights. Although neither of these models was developed with the needs of public science in mind, both provide helpful examples of how universities, federal funding agencies, and scientific bodies might contractually reconstruct a research commons for scientific data that could withstand the legal, economic, and technological pressures on the public domain.[7]

[7]See pages 425 to 456 of Reichman and Uhlir, op. cit., note 1. We propose the contractual regulation of government-funded data in two specific situations: (1) where government-funded, university-generated data are licensed to the private sector, and (2) where such data are made available to other universities for research purposes.

25

The Open-Source Paradigm and the Production of Scientific Information: A Future Vision and Implications for Developing Countries

Charles M. Schweik
University of Massachusetts, United States
J. Morgan Grove
U.S. Forest Service
Tom P. Evans
Indiana University, United States

This presentation focuses on how open-source programming and the more recent movement in open content licensing might provide the building blocks for a new paradigm in collaborative scientific research.

There are some challenges in communication that we all are aware of, including the space limitations of paper media; the lack of library resources, particularly in developing countries; the challenges of scientific repeatability; and, often, distributed and uncoordinated research efforts.

Peter Drucker made an observation about innovation lags that occur when new technologies are introduced.[1] In his paper he provides the example of a lag in railroad innovation in the United States in the nineteenth century. He states that railroads were first designed to move only people, and it took nearly 30 years before they were used to move freight. We would like to propose that in terms of the Internet, the Web, and the way scientific information is shared (e.g., current e-journals), we are currently experiencing such an innovation lag.

With the emergence of the Web the potential exists to innovate in terms of how scientific research is shared and Internet-based collaboration is accomplished. Instead of providing only papers with final results, the Web creates the opportunity to make available scientific content at various stages of the research process. Incentives for scientist participation are key, and therefore the idea of a peer-reviewed e-journal is vital. Consequently, we envision future Web-based e-journals as containing information at all stages of the research process. Systems of peer review could be instituted at each stage to ensure quality control and to provide an incentive for participation. This next-generation e-journal would also apply principles used in open-source programming collaboration, as well as open-content licensing, to promote distance learning between scientists and to speed up the innovation process. Such a collaborative paradigm could increase the ability to harness global collective action toward solving many difficult and complex problems that humanity faces and could have important implications for greater participation and communication between scientists and academics worldwide.

To describe this vision more fully this presentation will first discuss the general design principles of open-source projects based on research of literature on open-source programming, and some of our own preliminary empirical research. We will then discuss why we see the open-source model combined with the emerging phenomenon of open-content licensing as a potentially new paradigm for global collaborative scientific research. The

[1] P. F. Drucker. 1999. "Beyond the Information Revolution," *The Atlantic Monthly,* October. Available at http://www.theatlantic.com/issues/99oct/9910drucker.htm.

presentation closes with an example of our vision that we are trying to initiate: an open-source/open-content approach applied to land use/land cover change modeling.

PRINCIPLES OF OPEN-SOURCE PROJECTS

The basic principles of open source are grounded upon an innovation in software licensing. While there are a variety of open-source license types, in general they require the free distribution of the software coupled with readable source code. These licenses often allow new derivative works to be developed from the digital content (program source code). If you are a programmer, you can interpret the programming logic and have the opportunity to contribute new functionality to that software, with one catch: many open-source licenses (e.g., the GNU General Public License) are of a "viral" nature, requiring improvements to follow the same licensing agreement. The licensing in open source is a critical component in how these projects work.

The Linux operating system and the Apache Web server are two high-profile open-source success stories. They provide examples of how the open-source collaborative paradigm can produce solutions to very complex problems. The question is whether these are anomalies or whether there is really something to this collaborative approach that could be applied in other complex problem contexts.

To explore this we should understand better how these projects work. Based upon a review of recent literature,[2] a key attribute of Internet-based open-source collaboration projects is an established team of volunteer programmers and testers. In some cases (Linux being one prime example) there are organizations (e.g., IBM) that actually pay employees to work on the project.

Modularity and parallel development are also important design principles for open-source projects. The idea is that by dividing the project into modules or small compartments, people around the world could select a particular module to improve, without interfering with anyone else who is working on the project. Improved modules can be resubmitted back to the project, subject to a peer review. If the improvement is deemed useful, it is added to a future release of the software.

Typically open-source projects are initiated by an individual or a small group who have a critical need for software functionality that is currently unavailable or is too expensive to purchase. They then develop a "kernel," a core piece of software with some promise for potential innovation and growth. For a high-quality project to emerge highly skilled or prominent people in this particular area are required in order to give it promise for later participatory growth.

Internet-based collaborative infrastructure is also important for open-source projects to work. Here version control systems are crucial. These systems keep track of various changes to existing modules and allow the project administration to control and keep a record of different releases of the software. The Concurrent Versioning System is one of the most popular softwares of this type and is used in many open-source programming projects.

Project governance—including rules related to participation and conflict resolution and norms of behavior—is another potentially important aspect of open-source collaboration. To date there is very little literature addressing this subject,[3] but our hypothesis is that this is a very important factor that leads to the success or failure of these projects.

For many projects the goal is ultimately to achieve growth both in developer participation and in a user base. From a developer perspective this is really the law of numbers, where with more eyes problems are more easily solved.[4] The idea is to get a large participating group, hopefully globally, working on these problems. The high-profile success stories have achieved this; Linux and Apache enjoy a large community of developers with regional coordinators working in many languages. In the case of Linux there are approximately 18 different languages represented, and it is a potentially complex system of coordination and core staff.

---

[2]See C. M. Schweik and A. Semenov. 2003. "The Institutional Design of 'Open Source' Programming: Implications for Addressing Complex Public Policy and Management Problems," *First Monday* 8(1), at http://www.firstmonday.org/issues/issue8_1/schweik/.

[3]To examine some of the active open-source programming projects, visit http://sourceforge.net.

[4]E. Raymond. 1998. "The Cathedral and the Bazaar," *First Monday*, Vol. 3, No. 3, available at http://www.firstmonday.dk/issues/issue3_3/raymond/.

One aspect of open source that has puzzled scholars (particularly economists) is the question of incentives to participate. Why would people voluntarily and freely contribute their intellectual property to such an endeavor? Recent research has identified several reasons. The first is related to intrinsic motivations. Some people enjoy this work. They like programming and they find it interesting and creative to participate in these endeavors. Second, self-esteem sometimes plays a role. People often want to feel like they are part of a community and are contributing to some important endeavor. Third, there are altruistic and social and political motivations. The altruistic motivation for many open-source programmers is based on the idea that software should be free. Political motivations also can be important, as in Linux, where programmers participate in part to take on a perceived monopoly.

More recently analyses have shown that there are primary economic motivations, either to build skills for future economic gain or to self-promote. Regarding skill building, open source provides a valuable distance-learning function; over the Internet the programmer can look at the source code, learn about how other people approach a programming problem, try to do their own enhancements, and then participate in a system of peer review in which they receive critical feedback. This distance-learning attribute of open-source projects can be a significant motivator for participation. Self-promotion is another economic-related motivator. Becoming known in the community often leads to potential consulting or job opportunities.

Another motivation often reported in the open-source literature is personal need. There may be cases in which no software is readily available for a particular function or purpose. An individual programmer might try and work on the project but realizes it is a large or complex task. The programmer realizes he or she cannot do it alone and thereby embraces the open-source paradigm with the hope that someone else will share the work at hand.

This brings us to a very important point about the above summary of the characteristics of open-source projects: it is primarily based on available literature at the time of writing this presentation (February 2003), and this literature is simplistic and biased. One of the problems with the conceptualization of open-source projects as described above and its potential is that there is considerable hype surrounding it. This is largely because the literature to date focuses centrally on such high-growth success stories as Linux and Apache and ignores what could be thousands of failed projects.

There are many open-source projects that have been initiated, but of the more than 50,000 currently listed,[5] many never will achieve the level of success (in terms of growth of a participant base) of a Linux or an Apache Web Server. In fact it is highly likely that a large percentage of the thousands of open-source software projects are written by individual college students who decided to license it as open source and place it on the Web. To look at the open-source paradigm as a collaboration paradigm there is a critical need to study the factors that lead to successful cases (however it might be defined) and also unsuccessful (failed) cases. Our earlier discussion about the lack of research on the structure of open-source governance mechanisms is a case in point. There should be a major theoretical and systematic research endeavor (which is now beginning)[6] that examines these projects more deeply and identifies the factors and institutional designs that lead to successful high-growth or failed open-source collaborations. The fact that collaborations like Linux, Apache, and other open-source software exist and thrive suggests that there is something very interesting about the concepts of open-source (and -content) licensing, and the collaborative systems that have achieved them. We should understand better what factors lead to successful collaborations, not only so that other software might be developed in this manner but also and more importantly because it might provide a new way for humans to collaborate on a global scale on difficult scientific problems facing humanity; this approach could have important implications related to increasing participation from scientists in countries worldwide.

OPEN SOURCE/OPEN CONTENT AS A NEW PARADIGM FOR COLLABORATIVE GLOBAL RESEARCH

One could consider software development as one type of scientific endeavor. And one could argue that the principles of open source provide an opportunity to greatly increase the speed at which new innovation is achieved.

[5]For example, see http://sourceforge.net

[6]See, for example, the latest workshop on open-source software engineering, available at: http://opensource.ucc.ie/icse2003/.

There are several reasons for this. First, the incremental publishing feature allows much faster communication of new findings. Second, because of its open-access nature and because it is Internet based, the open-source paradigm has the potential of reaching a larger, potentially global audience. Third, the distance-learning attribute of the open-source approach provides an incentive for many to participate.

Let us expand a bit on this third point. In his book, *The Future of Ideas*,[7] Lawrence Lessig explains that the tremendous growth on the Web from 1994 to 2000 was based on two factors. The first is the end-to-end design of the Internet. The Internet as a communication system was designed with a relatively simple transmission protocol. The complexity is largely introduced at the end points, namely the end-user computers and servers. Programmers could generally rely on the fact that the transmission protocol will continue to follow the simple structure and design, thereby allowing data transmission through the network to be very simple and standardized. This allows them to place more sophisticated programming logic—the innovations—at the end points of the Internet (browsers and servers).

The second factor leading to the exponential growth of the Web is that the designers of Web browsers such as Netscape and Internet Explorer provided a "view source" option under their menus. In early days of the Web development (e.g., 1994 to 2000) the way people learned how to program Web pages was by visiting other people's Web pages with these browsers, viewing their source code using the "view source" option, examining what they did, and then developing their own Web forms based on that new knowledge. This leads us to some key points: While not acknowledged as such, the Web is arguably the largest and most successful distance-learning program in history, and it has led to tremendous innovation over the past six to eight years. Even through Web pages were not licensed specifically as open source, the "view source" option in browsers meant that the Web pages at this point were indeed source that was open access. The exponential growth in Web pages during this six-year period provides an important example of the kind of innovation that is possible by following an open-source approach.

Very recently the ideas of open-source licensing have moved into the broader domain of "open content." Essentially, open content extends the "copyleft" principles of open-source licensing into broader areas of any form of intellectual property. Any digital content, from music to academic papers, could be licensed in a similar fashion to that of open-source programs, where end users of the content are given permission to freely copy, distribute, and possibly derive new works based on the content. The Creative Commons project,[8] recently initiated by Lawrence Lessig and his colleagues, provides 11 different open-content licenses based on the answers of four different questions: (1) Is free copying and distribution permitted? (2) Is author attribution required? (3) Can derivative works be made from this content? (4) Can the content be used in commercial applications without permission? Different combinations of the answers to these questions create a spectrum of intellectual property rights, from full copyright on the one end to public domain on the other. By raising attention to this spectrum Lessig and colleagues show that authors of new content have more choice than just going with the traditional "copyright—all rights reserved" to "copyright—some rights reserved." For example, their "by attribution" license allows others to copy, distribute, display, and use copyrighted work, as well as produce new derivatives from this work, as long as they acknowledge the previous author. Adding their "no derivative works" license with "by attribution" means that people can copy, distribute, display, and use the work verbatim, but cannot define or develop new work using it.

Over the last year examples of what we refer to as open-content experiments have emerged. We refer to these as experiments because they have just started, and it is not clear which ones will actually succeed and which will fail. The oldest examples of this kind of idea include Project Gutenberg[9] (a public-domain project), which is placing e-books in the public domain on the Internet, the Free Software Foundation's GNU Free Documentation license[10] to support the open-content development of software documentation, and the Massachusetts Institute of Technology OpenCourseWare[11] project, in which academics share course content using open-content licensing.

[7]L. Lessig. 2001. *The Future of Ideas: Fate of the Commons in a Connected World*, Random House Publishing, New York.
[8]See http://creativecommons.org.
[9]See http://promo.net/pg/.
[10]See http://www.gnu.org/licenses/licenses.html#TOCFDL.
[11]See http://ocw.mit.edu/OcwWeb/index.htm.

OPEN SOURCE/OPEN CONTENT AND THE SCIENTIFIC ENDEAVOR: AN EXAMPLE IN THE CONTEXT OF LAND COVER CHANGE MODELING

We have been working on an interdisciplinary project studying how land cover is changing and how computer-based models can be created to capture land cover change dynamics. This is an issue of great worldwide interest. Local governments, for example, are interested in tools that might help them project future scenarios of sprawl and how various public policies might change such projections. The global change community wishes to use these kinds of models to understand how forests may change and what that might mean for the global climate system. The challenge in developing these models is that the system that drives land cover change is often extremely complex and involves insights from a variety of scientific disciplines. It is a problem that could involve a large number of participants across the globe, including academics, scientists, policy analysts, and local and regional governments.

In 2002 we conducted a review of existing land cover change models and identified more than twenty developed by various organizations and funded by a variety of agencies.[12] These models can be extremely complex to use and understand because of the different technologies and academic disciplines that are represented within them. Our ability to build on advances made in these models is relatively low, mainly because of the significant costs involved in acquiring needed software technologies, and learning and applying these models once the technical infrastructure is available. We see the open-source/open-content approach as a way to move beyond the status quo and possibly to speed up our ability to build on another's work much more rapidly than we have to date. Let us use this as an example of how we might apply open source to a collaborative, scientific endeavor.

To implement an open-source/open-content approach in this area the first step would be to identify a core group of willing participants. This would include modelers, who are similar to the programmers in open-source programming; data providers; scientists and academics; and other professionals who contribute theoretical arguments into these models, as well as the practitioners and other stakeholders who might be users and who also might be able to provide input into these models. The number and diversity of participants would be greater than in open-source programming.

The next step would be to identify existing models that could possibly be placed under an open-source/open-content license. In this context we would extend the idea of the open-source software "kernel" to land use models, making them available with an open-source license. Unlike traditional programming, in this scientific endeavor there would be multiple kernel types. The model itself (in whatever technology or approach) would be one type of kernel. These models are often informed by theoretical work and this content is often in the form of published or unpublished papers. These "theory kernels" could be placed under an open-content license. Data required to run the models could be considered a third type of kernel that could also fall under an open-content license. In the initiation phase all these land use modeling components or kernels would be modular to support parallel development.

An important consideration at the initiation of an open-source/open-content based scientific endeavor is the incentive structure to encourage participation by scientists worldwide. Interestingly the motivations of programmers in open-source projects and the motivations of scientific and academic researchers are very similar. They share the same intrinsic motivations. Researchers often participate in the creative process because they enjoy it or find the subject interesting or important to contribute intellectual property to it. There can be a self-esteem component that motivates their participation as well, a feeling that they are participating in a broader community of interested scholars. In terms of altruistic and social political motivations, researchers are often motivated in their work because they are trying to solve a problem facing humanity. There is also the belief among many researchers that, like software, knowledge should be free, which might motivate them to participate in an open-content project. As such, there is a potential social and political movement very similar to the free software argument. In terms of the personal needs motivation that drives many open-source programmers, instead of a software gap there is often

[12] C. Agarwal, G. M. Green, J. M. Grove, T. P. Evans, and C.M. Schweik. 2002. *A review and assessment of land-use change models: Dynamics of space, time, and human choice*, Gen. Tech. Rep. NE-297, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA., p. 61. Available at: http://www.srs.fs.usda.gov/pubs/viewpub.jsp?index=5027.

a scientific knowledge gap. The knowledge gap is complicated and often requires multiple disciplines, and researchers find themselves needing the assistance of others with other expertise.

The motivations described above in part could encourage scientists to participate in an open-source/open-content research effort. But as in open-source programming, we believe the most significant motivation is along the economic dimension—to gain new skills and to self-promote one's ideas. Researchers might participate in an open-source/open-content scientific collaboration to building new skills. For example, the global body of graduate students interested in land use change issues would be motivated to participate in such a project because of the benefits of what they could learn. Having research products as "open" provides the ability for one to distance learn the way scientists already distance learn by reading academic journals. Open-source/open-content collaboration would provide a setting for peer review as in open-source programming. Peer review is an important way that scientists learn from each other and it is a way for us to make advances in our own skills and knowledge. Self-promotion, or becoming known in the field that you are studying, is a motivation that is particularly important for academics, and especially for junior faculty who are trying to make a name for themselves. Graduate students could possibly gain some real self-promotion if they contributed an important new module (e.g., theory or computer code) to the effort. In short, the motivations driving potential participation in open-content scientific collaboration are very similar to what motivates some open-source programmers to participate in those projects.

One of the great challenges to the open-source/open-content vision applied to the scientific endeavor is the traditional way scientists are promoted through peer-reviewed publishing. Junior (and higher-level) researchers in university settings are often evaluated on their publishing record and therefore have a strong motivation to protect their intellectual property accepted through traditional publishing media. The concept of a next-generation peer-reviewed e-journal is important, because it considers these important incentives for participation. What we are suggesting is that these next-generation e-journals provide mechanisms to "publish" various forms of intellectual property. In our example of land cover change modeling this might include papers on final results, but also papers on theoretical inputs that inform a model design, papers on how to use the model, as well as the other important products for the modeling endeavor, such as data sets and the model source or logic itself. If a researcher contributed a new module to the model (e.g., extended the model's functionality), this would be peer reviewed and considered a publication as well.

What this means is that there is an important need to figure out how to document intellectual property contributions for all three of those kernels—models, data, and the theoretical contributions. Can all these components of the research process be treated as a form of publishing, or possibly service contributions? This means that we will have to establish a system of governance and rules of operation for this next-generation e-journal (e.g., editorial boards for all kernel types), as well as effective conflict resolution mechanisms to govern a debate about which direction the modeling process should go.[13]

Creating the next-generation e-journal would also involve the selection of appropriate open-content licenses for these kernel types and the establishment of systems of peer review—not only for papers, which may be the theory or empirical work, but also for data and models. The final component would be to establish a project infrastructure, which would include the communication and version control systems.

Just as in the case of today's e-journals, theory, empirical research, and results would be critical submissions. In the next-generation e-journal we could see following the model of traditional volumes and issues, and also the ability to produce incremental releases. In the electronic world there is no necessity to "print" a particular issue, rather you can just build it as new contributions are submitted and peer reviewed. Moreover, in land cover change modeling the opportunity exists to make hyperlink connections between related kernels. For example, links could be implemented that make it easy to see which data feed a certain model, or which theory (papers) informed a particular modeling logic.

We are associated with several major research groups interested in land cover change modeling, including the U.S. National Science Foundation's Long-term Ecological Research Network and Human Dimensions of Global

[13]Conflict resolution mechanisms are not discussed in the open-source literature. For example, there must be conflicts between two modules that are competing with one another; how is it decided which goes into the next release of software?

Change research network. Along with the U.S. Department of Agriculture Forest Service we are trying to develop a core group of people interested in starting open-source/open-content collaborative research effort on land cover change modeling. At the time of this writing we have established a working group and have held a workshop to initiate the endeavor. Our short-term goal is to develop an institutional design around the collaboration, including the identification of available models, establishment of metadata standards for data and model modules, identification of the types of open-source/open-content licenses for various kernels, and the establishment of required communication systems and version control systems. Over the longer term we would like to move to a next-generation e-journal, perhaps in collaboration with an existing e-journal.

CONCLUSION

Several conclusions can be made from this discussion. First, there is a great need for more in-depth research on open-source programming and how these projects work if we want to capitalize on them in the broader context of open content. The literature up until early 2003 is fairly naive on how open-source projects are structured, and a major point of our talk was that there is a desperate need to understand what leads to success and failure of these projects, rather than basing all our knowledge on just the high-profile success stories, Linux and Apache Web Server, which may be anomalies. Fortunately more carefully crafted quantitative studies and deeper analyses on these projects are beginning to emerge.

Second, we should watch the open-content experiments and understand which ones are successful. There is, as far as we can see, little research on how these projects are managed. We are trying to initiate this kind of research.

Third, in scholarly (global) communication we should move beyond Drucker's innovation lag in current e-journals that follow the old volume-and-issue paper model and encourage the development of the next generation of e-journals that takes advantage of all the Web can offer. We are suggesting that convergence with the open-content licensing phenomenon might be a component of this, and that the idea of not only peer-reviewed papers but also data and other computer-based scientific research (e.g., computer models) will be important. This lag in scientific publication is similar to the current e-commerce and e-government movements, which have been gradually moving from simple Internet publishing of information to more sophisticated online transaction processing.

Our major point is that the continued growth of access to the Internet globally, coupled with the design of open-source licensing and collaborative principles and the emerging trend in open-content licensing point to a new, potentially significant way for humanity to tackle complex problems in science and other domains. There is still much to be learned about what works and what does not in open-source-like collaboration. But the promise is there; successful global, Internet-based collaboration of complex problems in open-source programming (e.g., Linux and Apache) suggest that there is a way to achieve global collective action in ways never before possible. It may lead to no less than a paradigm shift in the way new scientific knowledge can be generated and the way new learning is shared. Assuming Internet infrastructure continues to be built in countries that lack access, this open-content (e.g., open access), next-generation e-journal we envision and open-content collaborative projects could dramatically improve the sharing of knowledge and participation in all corners of the world.

26

New and Changing Scientific Publication Practices Due to Open-Access Publication Initiatives

Erik Sandewall
Linköping University, Sweden

During this symposium there have been a number of examples and references to legal rules that apply to research activities and economic rules and conditions that influence our lives as researchers. There are a number of other rules that are equally important for us in our professional life, but which are internal to the scientific community. These are the rules and practices that govern publication activity, at least in the broad sense of publication.

THE PRINCIPLES OF SCIENTIFIC PUBLICATION

There are several principles that have guided the publication of traditional scientific journal articles, which include the following:

- In science, technology, and medicine, quality control by peer review is key. It is highly inappropriate to publish an article describing the results of a colleague. It is understood that one should not, even though such an act would not constitute a copyright violation.
- A journal will not republish a previously published article, even if the copyright situation is correct. This was a principle introduced by the *New England Journal of Medicine* in the early 1950s, and it has been generally embraced.
- Correct reference should be made to the first reported results. The priority of results is an important part of the incentive system for researchers.
- Reviewers are anonymous in almost all journals.
- Reviewers are supposed to be objective and not favor or disfavor a paper simply because it relates in some way to their own interests. Papers are considered confidential while reviewed.
- Priority is counted from the day of publication of an article, that is, a paper must be reviewed before priority can be given.
- Institutions have differing rules, such as whether advisors should be listed as an author of a paper published by their students.

Interestingly, these principles are not in general the results of legislation and they have relatively little connection to the legal system that constrains researchers. For example, the copyright rule would certainly prevent

a researcher from publishing verbatim an article previously published by someone else, but it does not prevent someone from republishing the previous results in their own words. Economic mechanisms are irrelevant. What we have are social rules within the research community, which are in part policies that have been established by journals, universities, or other organizations within the community. This set of rules is specific to the research community. The legal system oversees the community but not actively. There also is a system of incentives in the research community that seemingly contradict economists' claims that incentives should be monetary because people are driven by monetary compensation. These economic claims are simply not true.

Thus, rules observed in economics that apply to physical property do not automatically transfer to intellectual property. On the other hand, there are other economic rules and results that are more appropriate for the scientific community.

For example, in the 1930s this economic question arose: if the market is such a wonderful design, why do enterprises of more than one employee exist? How come everyone has not formed their own company? The answer, from a purely economic view, is that the transaction costs would be too large. So an enterprise is organized with internal incentives to keep transaction costs down. Certainly in science today when we see enormous transaction costs, for example, for publication and risk communication, we should consider this economic model.

There are two reasons for economic rules that govern the scientific community: (1) efficient dissemination and preservation of scientific information and (2) provision of efficient and varied incentives. Finally, let us observe that these kinds of rules and practices are highly technology dependent. For example, the current peer-review system requires that articles after submission must be sent to reviewers in a different part of the world for review.

Could such a review system have existed before current technology? It would have been very difficult. When there was no photocopying, and handwritten manuscripts were given directly to typesetters, it was very difficult to obtain several copies of a manuscript. Modern peer review became possible with the advent of typewriters and carbon paper.

Now, of course, we are experiencing another enormous wave of technological changes. In fact, the very possibility of open access is due to the technological revolution in the 1990s. We should ask the question whether it will again be necessary to revise current rules of practice because of technological change.

CHANGING PUBLICATION PRACTICES

The rule about previous publication was established by the *New England Journal of Medicine* before the advent of the preprint, mimeography, and widespread photocopying. When preprints emerged a few years later, there were two different reactions in research communities. Some people believed that under this rule, papers that were presented as preprints could not be published. This meant that you could not use preprints. Other communities believed that a preprint was not really a publication, which is an astounding interpretation of the word "publication" from the patent lawyer's point of view or just from ordinary common sense. This strange use of "publication" has persisted even to today.

The moral of the first example is that as technology advances rules regarding publications may need revisions. In another, more recent example one major publisher removed a number of research articles from listservs in response to pressure from some groups who considered the content or the wording inappropriate. If you only carried the electronic subscription, you lost access to the articles. The scientific community has, of course, serious concerns as to whether this is reasonable. This was not technically possible before. A policy must be established to govern electronic publications under these conditions.

The third example occurred in 1997 when I started an electronic journal and experimented with another system implementing a two-stage review process. The first stage consisted of a three-month open discussion period in which papers were posted on the Internet and peers were invited to submit comments. It was not anonymous; rather it was a discussion just like in a conference. After three months authors could revise their paper, which was then sent to confidential pass-or-fail referees. This scheme had many advantages: more safety, fairer treatment of the authors, better rewards for the reviewers, and better political control of reviews.

One concern that we immediately encountered was that during the period of open discussion, someone might steal the results, publish the paper in his or her own name, and get the priority. The only solution was to decide that priority begins with the date of the first appearance of the paper in discussion, that is, publication counts before the reviewing begins. First you publish, then you review, and then the journal, if the paper is accepted, technically republishes the paper. In order to use the more advantageous review process it was necessary to change the mode of thinking and the terminology.

Another example is the placement of new research results directly in a database without any article to document the results. We should have a way to characterize a database contribution as an entity on a par with a published research article, because data that are in the database will be used and later work will rely on earlier work. This imposes technical requirements on those databases. Certainly, when this happens on a large scale the scientific database system must keep track of those references, as publishers keep list of references for an ordinary article. The question is should that reference to other work be presented and if so, when and how, and if it is not in a database, what other mechanism can the scientific community use to give incentives for such work.

We ought to think of scientific publication as a function that is internal to the scientific community. With some exceptions, such as in medicine, papers are published by scientists for scientists. But the scientific community has chosen to outsource this function to mainly commercial actors. By doing so we have established contracts with commercial partners that seem to be enormously favorable to them in economic terms. We certainly should not complain, however, that the commercial publishers earn money, because it is their objective to make a profit.

Instead, we might think once again about the agreement we have negotiated. We should be a bit smarter about the deals we establish when we outsource functions that are important to us. This might also require us to look over the rules and conventions that we are using. Those rules and conventions primarily should serve our interests as researchers and make the research process work well, but there are also pragmatic aspects. For example, consider an editorial board of a journal that is independent and uses a commercial publisher. The board is dissatisfied with the pricing policy of the journal, which they consider too expensive. They decide to solve the problem by establishing a new journal with another publisher. They cannot bring the name of their old journal with them, because the name of the journal is owned by the publisher; so they create a new name. The entire editorial board resigns and goes over to the new publisher to start the new journal. The old publisher, then, recruits a new editorial committee.

An interesting question arises: should the impact factor and the prestige be ascribed to the name of the journal or to the editorial board? There are two reactions. One might naturally think that in this situation the prestige is derived from the name of the journal. The other reaction is to let the research community decide, since impact factors and prestige are used by tenure committees, granting agencies, and other organizations within the scientific community. We could certainly say that in this case the impact factor goes with the editorial board, but it will require action. Currently the impact factor stays with the name, making it difficult to start new journals. This is one of the reasons why it is difficult to do what this editorial board did, and it is one of the reasons why the people on that editorial board now consider what they did a bad idea.

One result of new technology that is evident on a large scale is the increasing do-it-yourself activity and self-publication, either by the individual researcher, a university, or more centralized archives. In the case of university publication this sometimes falls under the term of universal electronic press. As a result two issues that are important to the scientific community arise. The first is that the long-term preservation of the publications must be guaranteed. It is not appropriate for publications to vanish within two years.

Second, it is also very important that no one is able to manipulate the articles after publication. In particular the authors should certainly not be able to go back and improve their articles two years later while keeping the old time stamp. This fact must be absolutely clear to the rest of the world.

Therefore, there must be control schemes. You might imagine several different control schemes, for example, duplication of multiple copies of the paper in different parts of the world or an encryption scheme with public keys and so on. It is important that these schemes require public awareness and public acceptance. The scientific community must consider this issue when establishing rules and procedures.

The last example relates to Robin Cowan's illustration[1] of the book containing a table of integrals that is no longer commercially viable because it is on the Internet. Resources that existed before a new technology and were sold to researchers at a reasonable profit may no longer be viable, just as horse-drawn carriages have become obsolete. If this occurs, the incentives provided within the scientific community may have to be broadened to give better incentives for things we previously did not give incentives to because they were taken for granted.

Today in science we tend to give much more credit for doing original or new work. We should remember, however, the nineteenth-century work of making digests, summaries, surveys, and compilations of earlier work is still very important for the research community. Perhaps more credit should be given to these kinds of ventures.

CONCLUSION

Publication policies tend to be long lived as illustrated in the example of the *New England Journal of Medicine*, which established a policy in 1951 that became less relevant in 1960.

When we discuss policies today, especially at this time of rapid technological development, we should try to anticipate the development that is likely to occur in the future and how our policies will relate to it.

I suggest that the direction we are going is research knowledge management. In research there are documents of different kinds—articles, research reports, reviews, discussion items, raw data, and laboratory notes from experiments. This broad collection of information is not organized in a very coherent way. Some organizations have begun to address this issue.

There was another conference in Paris in January 2003 where one major French research institute described its plans for building exactly this kind of system for its own needs.[2] This is a very likely development. It will continue on toward integration on the international level within science, integration or interfacing with knowledge management systems in industry, and integration or interfacing with knowledge management systems in government or the public sector. This trend may be quite relevant to the scientific contribution at the World Summit on the Information Society in Geneva. Thus, when we review our own policies and rules for scientific publication and communication, we ought to keep this in mind.

A change in the direction of such knowledge management is only going to be acceptable and useful if the research community is able to revise the rules so that the incentives and values important to us will function well with the new technology.

Is it even possible to change the existing rules and practices? It is not easy. Nobody can dictate it. That change certainly must arise in the same way that those rules and practices arose in the first place, that is, by discussion within the scientific community that leads to new policy decisions by such organizations as journals.

What is needed is a broad discussion by a body such as the International Council for Science (ICSU). These issues must be brought to attention and the discussion ignited, perhaps with provocative proposals.

The Committee on the Dissemination of Scientific Information within ICSU might be a group that could initiate such debate. The mandate of this group is to do studies that lead to quality proposals for ICSU. Publication policy is a natural item for the agenda for this committee.

---

[1]See Chapter 8 of these *Proceedings*, "Economic Overview of Open Access and the Public Domain in Digital Scientific and Technical Information," by Robin Cowan.

[2]For more information, see the International Council of Scientific and Technical Information (ICSTI) Open Access Meeting Web site at http://www.icsti.org/open_access2003/index.html. This meeting was convened by ICSTI, in partnership with the INSERM and the French Ministry of Research, and in association with ICSU and the Committee of Data for Science and Technology.

27

Overview of Open-Access and Public-Commons Initiatives in the United States

Harlan Onsrud
University of Maine, United States

This presentation provides some examples of open-access and public-commons initiatives drawn primarily from the United States. To place the presentation in context it should be noted that most wage earners in the world can ill afford the laptop computers with which we have been making our presentations at this symposium. For well over half the nations of the world the cost of this laptop computer exceeds the entire annual salary of the typical worker.

In contrast, most citizens in the United States and in much of Europe can afford and have access to a computer. Many families own several computers and practical access to high-speed cable networks is now almost universal in the United States. The average U.S. or European citizen is far more able to connect with others in contributing to digital collaborative projects than are average citizens in much of the rest of the world. However, the incentives to work collaboratively toward open-access goals may not be as great when you can afford to buy. Incentives to collaborate also are not as great when you do not have a clear vision of how you could productively contribute to a shared open-access depository or development effort.

Despite the impediments large numbers of people across the globe are expending considerable effort in developing open-access and open-source resources. An open-access or "public-commons" approach to resource development can be defined as an end product that is free for anyone to access, utilize, copy, and make derivative products although some limited restrictions may be imposed in order to enhance retention or distribution of the resource's public availability.

Open-source code and open-access models are sometimes viewed as alternative economic models or a new mode of production in which individual contributors are organized neither in response to price signals nor by firm managers.[1] Under certain conditions this new form of production makes sense and works, while under other conditions it does not.

Some people view these new cooperative means of production as a supplement, complement, or replacement to government funding for the production of public goods (i.e., goods that are nonrival/nondepletable and nonexcludable/nonappropriable). That is, these are the types of goods that will not be produced by normal marketplace dynamics and therefore, if they are desired by society, must be produced through some other means.

[1]Yochai Benkler. 2002. *Coase's Penguin, or, Linux and the Nature of the Firm*, *The Yale Law Journal* 112, winter 2002-2003. Available at http://www.yale.edu/yalelj/112/BenklerWEB.pdf.

In some instances it makes far more sense to produce such information through direct government funding. In other instances public commons approaches perhaps can be more efficient or complement the government provisioning efforts.

Other people view organized open-access and public-commons approaches as systematic means for organizing volunteerism and philanthropic efforts for the production of communal information products and services. Regardless of theory one can obviously make information fully available to others without one's own self subscribing to any of the explanatory hypotheses.

This presentation reviews illustrative open-access and public-commons projects under five general headings: (1) open-source software; (2) open-access journal and article initiatives; (3) open-access disciplinary and institutional depositories; (4) search engine approaches; and (5) the general public-commons types of initiatives.

OPEN-SOURCE SOFTWARE

The vast majority of speakers at this symposium have used the leading commercial presentation software to guide their talks. Yet, the freely available open-source equivalents of Microsoft Office have increased dramatically in quality, even within the past few months. For example, OpenOffice is an open-source application that extends from and builds upon the open-source operating system of Linux.[2] This particular open-source suite of programs seems to work well, the programs are free, and many argue that the potential exists for these programs to far surpass the Microsoft Office suite in general quality and usability due to the transparency of the code and the high level of private and corporate commitments to continue to improve the programs.

Who wrote the code? Volunteer programmers continue to contribute, but this version is also strongly supported by Sun Microsystems. Sun sees at least some open-source software as making sense for its long-term corporate business model.

Writing computer code for sophisticated products such as Linux or OpenOffice requires a high level of expertise. The typical person and even the typical programmer cannot productively contribute to such a development effort. Often these high-end products will take much more than volunteer efforts to make them truly useful and practical. Regardless, the resulting software becomes available for all in the world to freely utilize.

JOURNAL AND ARTICLE OPEN-ACCESS INITIATIVES

There now are several major specialty collections of full-text, open-access scientific journal articles freely available on the Web. For example, the National Aeronautics and Space Administration's Astrophysics Data System has 300,000 full-text articles online[3] and Highwire Press has about the same number focused in the biomedicine and life science fields.[4] Other initiatives include the high-energy physics arXiv[5] and PubMed Central.[6] Most of these online archives deal with intellectual property issues on a journal-by-journal negotiation basis or have scientists submit original work directly to their archive.

The U.S. National Library of Medicine (NLM) subscribes to many thousands of journals. NLM facilitates open access to approximately 100 life science journals that have agreed to make the full text of their articles freely available as soon as they are published, or after a specified period of time, through PubMed Central. Current PubMed Central journals have delays of up to two years, with most releasing their material six months or less after publication. Thus, one could argue that this open-access capability is funded by the philanthropy of those journals choosing to participate; however, several have reported that their subscriptions have increased since joining PubMed Central due to the visibility created.

---

[2]See http://www.openoffice.org.
[3]See http://adswww.harvard.edu.
[4]See http://highwire.stanford.edu.
[5]See http://www.arXiv.org.
[6]See http://www.pubmedcentral.nih.gov.

Elsewhere in the world BioMed Central supports 70 online fully-refereed medical and biology journals.[7] These articles are immediately open to worldwide public access. BioMed Central's economic model for providing open access is to charge author fees and fees from their institutional members. In this way the costs are spread among the numerous funding agencies to which authors pass on their charges.

The Public Library of Science is another initiative focused on the life sciences and medicine.[8] These are fields in which substantial research funding is currently available. Their economic model is very similar to BioMed Central, and they are contemplating a $1,500 author charge per article to support the system. Many of the world's open-access journals are accessible through the Directory of Open Access Journals.[9]

But, if you are engaged in scholarly research and are writing in the humanities or social sciences, fields in which government funding for most research is not the norm, the BioMed Central model is unlikely to work. In these scholarly areas, you cannot pass publication fees on to funding agencies. Or perhaps you are a young researcher lacking the funds to support your publication costs or you are economically disadvantaged in some other manner. Many researchers in the humanities and social sciences will be unable to contribute to pay the author charges that would allow them to contribute to the open-access scientific literature stream.

Another successful model works through professional member organizations. Dues are paid to a professional member organization that publishes a journal or database they make freely available through an open-access environment. Some free rider problems still exist, but this is generally a successful model for many journals.

What about researchers and scholars from developing countries? Again, many will be unable to pay the author charges that would allow them to contribute to the scientific literature stream. Even when these researchers gain Internet access and are able to read online journals, they are unable to contribute back to that journal under an economic model that requires the contributor to pay.

There is no single model appropriate for all fields and researchers. Other open-access funding models are working in providing common bodies of scientific literature open to all. For example, the Scholarly Publishing and Academic Resources Coalition (SPARC)[10] is an initiative of university libraries and others that have come together to pool their economic resources to start up and support low-cost journals that will compete directly with the most expensive academic journals. All libraries commit to subscribing to the new journals so that they have an assured minimum income stream that they can count on. Unfortunately, SPARC has very limited resources and therefore has had to focus on the highest payoff opportunities. As a result SPARC has failed to spin off new journals at the same rate as the private sector. SPARC recognizes this and has developed a revised strategy to provide a leadership role in exploring academic alternatives for supporting open-access publications.

Another approach is to develop open-access depositories for articles and data sets. Authors can avoid copyright problems with private publishers by openly archiving an electronic copy of their article prior to submitting it to a publisher for peer review. Most private publishers still agree to publish this openly archived work, knowing they are unable to require its removal from the depository. This is a prevalent practice; authors publish on their own Web pages or in more centralized depositories. Thus, open self-archiving of preprints is being highly encouraged by many in the open-access community.

Institutions setting up these archives tend to establish depositories that comply with the Open Archives Initiative using open-source software programs. Examples of such open-access depositories include the Eprints depository at Southhampton University, the DSpace repository at the Massachusetts Institute of Technology, and CDSWare at CERN.[11]

The only cost of setting up a preprint service at a typical university is technician time and server time. Article submissions are executed by the academic authors and handled automatically. Depositories can be set up to serve

[7]See http://www.biomedcentral.com.
[8]See http://www.plos.org.
[9]See http://www.doaj.org.
[10]See http://www.arl.org/sparc.
[11]Information on these projects can be found on their respective Web sites at http://eprints.org, http://www.dspace.org, and http://cdsware.cern.ch.

a specific discipline, such as the arXiv server. They can be set up to contain all publications arising from all disciplines at a specific institution, such as at a university. The Eprint site lists 66 organizations that are running archives with its software.

What are the drawbacks of the open-access, preprint depositories? They are only half of the solution. After an article is peer reviewed and published the author has to go back, update the metadata in the system, and file a corrigendum showing the changes to the article. The author cannot legally file the peer-reviewed article. While these systems are fine for preprints, almost no one does the coorigendum and metadata updates. Even if someone does, the article still is not in the form in which the author would like it to be read.

SEARCH ENGINE APPROACHES

Search engine approaches exist to provide full-text access, but they have a few problems. CiteSeer is one approach that is being used to index and access the computer science literature. It searches and crawls the entire Web.[12] The system uses an algorithmic approach to find citations that are germane to computer science literature, providing direct links to any full-text article that is found. It works on a citation-to-citation basis. The CiteSeer Index Web site has approximately 5 million distinct citations within computer science literature that have been drawn from about 500,000 full-text online articles. This is purportedly the largest full-text collection of scientific literature on the Web.

The legal problem with this approach is in obtaining permissions to copy the 500,000 articles. The system should automatically copy the journal articles in order to test the article against profile conditions, extract and index the citations, and host copies of the full-text PDF or Postscript files. The system developers have taken the position that they gain the substantial legal protections granted to search engines by the U.S. Digital Millennium Copyright Act (DMCA). Thus, crawlers like CiteSeer, Google, and AltaVista are able to cache copies of articles in order to allow their search services to operate more efficiently.

While CiteSeer began by crawling the Web, the vast majority of the URLs the system searches today are submitted by the authors who have posted their full-text and often fully refereed articles on the Web. The DMCA indicates that the crawler host is not legally liable if someone else, such as an author, submits an article in which that submitting author no longer has the copyright. So far publishers are not suing scientific authors for posting their own articles on their own Web sites, so these types of systems are actually working.

GENERAL PUBLIC-COMMONS APPROACHES

For those uncomfortable with walking a legal tight rope, the Creative Commons project offers some hope.[13] This is a thoughtful project, and once some further technical problems are solved, the approach could be embedded on the front end of every open-access data archiving and literature archiving project.

Creative Commons provides an online licensing approach that facilitates the ability of authors and artists to affirmatively place their works into the public domain or into a public commons legal environment. The approach can be applied potentially to all works, whether music, literature, databases, videos, or digital art.

How does it work? One goes to the Creative Commons site and chooses the restrictions, if any, that one wants to apply to their creation. The system automatically generates the specific open-access license to be applied to the work—one version is in plain language, another is in language that only a lawyer could understand, and a third version is machine readable to facilitate searching across the Web. When authors post their work on a Web site for others to download, they can also include a bit of HTML code automatically supplied by the Creative Commons system. When people click on the Creative Commons icon on the author's Web site, the restrictions chosen on that specific work are readily made known. A link is also provided back to the Creative Commons Web site for the full license provisions.

---

[12]See http://citeseer.com.
[13]See http://creativecommons.org.

What are the drawbacks of the system? The process currently applies only to the Web. Once a licensed file is downloaded, the user may have no idea which of the standard provisions apply since they do not accompany the file. Exploring technological solutions for ensuring that the identity of any file in any format is maintained and thereby retained in the public commons is a major development thrust of the project at the current time.

A second drawback is that finding files licensed under a Creative Commons license currently is very difficult. The team has talked to Google with the idea that a researcher will eventually be able to do searches on Google that are limited to public-domain and public-commons works. With this capability, a user will have at least some confidence that what they have found through this process, whether a music file or journal article, can be used freely without breaching someone's copyright.

A third drawback may arise if the Creative Commons Project proves to be a great success and hundreds of scientists start attaching open-access licenses to their articles and data sets before submitting them for peer review. Some scientists already attach such licenses to their submitted journal articles. Those articles are summarily rejected by most publishers without even being subjected to peer review. Will increased submissions by other scientists help place pressure on the publishers? Possibly, however my experience is that most scientists are likely to buckle. The current scholarly reward system is such that most scientists are more concerned with whether the journals in which they publish are ranked in Science Citation Index or Social Sciences Citation Index than whether they are broadly accessible. The reward system is not focused on the bigger picture of overall progress in science. We should change the reward system for the individual scientist decision maker.

THE IDEAL OPERATIONAL ENVIRONMENT FOR ACCESSING SCIENTIFIC LITERATURE AND RESEARCH DATA ACROSS THE GLOBE

Most researchers want the ability to cite across any and all scholarly domains and link from any citation found on the Web to the full article or the full data set on the open Web. That is what open access is all about; we would like to be able to use the Web as one large open library for us to share with one another. Open-access electronic journals are not likely to completely replace the commercial scientific literature, but open-access literature has a potential major role to play. Most researchers realize the benefits of having access and freely available access to one another's works.

The secret to open access, according to Peter Suber, is to keep control in the hands of those who most want open access—the authoring scholars. How do we keep control in the hands of the authoring scholars? How do we affect the decision making of individual scholars so that they retain power over their articles? There are several practical actions that can be taken to change the reward system. We should, for instance, consider changing the policies of funding agencies. These policies should encourage researchers to report in their grant applications only those articles and data sets that are in open-access archives. It does little good for a reviewer to assess another scholar's work or research proposals unless the reviewer has access to all the relevant significant works created by that other scientist. The current system of limited access for scientists in other than the wealthiest of institutions supports lost opportunities in advancing the progress of science.

We should be changing promotion and tenure policies. Peer-reviewed data sets and articles placed in open archives are much more valuable to society, and therefore ought to be recognized as such. The work of university scientists should be available to the world and not just to a small population of economically privileged scientists.

We should also change university intellectual property policies. Formal university policies should encourage professors and researchers to use open-access licenses and should give them full authority to use such licenses for their intellectual property.

Finally, we should identify within each of our disciplinary domains those journals willing to accept open-access licenses and those that are not. We should identify those journals allowing authors to post final journal articles on the Web and those that are not. The goal is that the reward system will eventually benefit economically those that follow open-access approaches.

If the reward system for scholars is restructured and online facilities were made easy, would individual scholars across the globe make use of open-access methods and archives to make their works available for sharing with others? I believe that the history of science shows that the majority of scientists would do so.

SESSION 6: EXAMPLES OF NEW INITIATIVES IN DEVELOPING COUNTRIES

28

Introductory Remarks by Session Chair

Alexei Gvishiani
United Institute of Physics of the Earth, Russian Academy of Sciences

This session focuses on new initiatives in developing countries. In addition to the projects highlighted by the speakers, there are several other projects that have been developed within CODATA, in collaboration with the Earth Data Network for Education and Scientific Exchange (EDNES), to transfer and promote new technologies to developing nations. These three projects are being conducted in countries of the former Soviet Union, which are now commonly referred to as the Commonwealth of Independent States (CIS).

In 1996, thanks to UNESCO, the STACCIS project (Support for Telematics Applications Cooperation with the CIS) was initiated. The goal of this project was to promote and adopt European-developed informatics technologies in seven CIS countries. The source of funding is continuing through the European Commission. The STACCIS project is quite successful; it received the highest grade at the European Commission evaluation in 1999. This initiative was extended and three projects were submitted and accepted by the European Commission.

Of those three the first project is Telesol,[1] a collaboration between the United Nations Industrial Development Organization (UNIDO) and EDNES, with the active participation of CODATA. Telesol promotes telework collaboration in business and research between the European Union and Central Asian, Caucasian, and Slavic countries of the CIS. A similar project is Telebalt,[2] which uses teleworking as a tool for information society technologies in Latvia, Lithuania, and Estonia. The third project, WISTCIS,[3] focuses on new methods for promoting information society technologies in the CIS.

In all three projects the main tools are conferences and training workshops in which the new technological products are brought to a country and local participants are taught how to use the products in the most efficient way. Another important tool is representation on the Web, and the different types of portals and Web sites that are equipped with modern tools, such as virtual presence system tool kits and collaborative browsing tool kits.

These activities are focused on open access to data, specifically research on environmental matters. Similar projects could be formed and submitted to relevant funding agencies for other parts of the world, such as Asia, Africa, or South America. These activities are quite useful in promoting modern technologies that deal with data.

[1] See www.ednes.org/telesol.
[2] See www.ednes.org/telebalt.
[3] See www.ednes.org/wistcis.

29

Overview of Initiatives in the Developing World

Sarah Durrant
International Network for the Availability of Scientific Publications, United Kingdom

The International Network for the Availability of Scientific Publications (INASP), based in Oxford, England, is a program of the International Council for Science (ICSU). INASP was established in 1992 following the ICSU/UNESCO Conference on Access to Information in Developing Countries. There are programs and activities in over 120 countries, principally in Africa, Latin America, and Asia, as well as newly independent states. INASP's mandate is to improve worldwide access to information and knowledge through a commitment to capacity building in developing and transitional countries.

OPEN ACCESS IN DEVELOPING COUNTRIES: CONTEXT, ISSUES, AND CHALLENGES

In 1996 Richard Horton of *The Lancet* talked about mainstream publishing condemning most third-world efforts to invisibility. He talked about thwarted efforts, and hindered development in developing countries.

> The invisibility to which mainstream publishing condemns most third world research thwarts the efforts of poor countries to strengthen their indigenous science journals, and with them the quality of research, in regions that most need them.[1]

The knowledge gap, or digital divide, runs in three directions: South to North, North to South, and South to South. In the South-to-North divide it is well documented that there are limited local resources in developing countries, limited information communication technology (ICT) skills, and a lack of access to technology. These are familiar challenges. In some quarters there is a perceived lack of credibility that we should overcome. Publications in the sciences in the South are in some cases not valued as highly as those originating in the North.

In the North-to-South flow information has become a very expensive commodity, and developing countries struggle to afford commodities. There are also digital divide issues that require technology and training. Initiatives such as INASP's Programme for the Enhancement of Research Information (PERI) and Health InterNetwork Access to Research Initiative (HINARI) of the World Health Organization, and also in a specialist field the Ptolemy project from the University of Ontario, enhance the flow of information from the North to the South.

PERI works in four related areas to support equitable access to and dissemination of learned information. These components will be described in more detail later on in this paper.

---

[1]Richard Horton, ed., *The Lancet*, 1996. See http://www.ias.ac.in/epubworkshop/presentations/epub_DCs/tsld006.htm.

HINARI was created to bridge the "digital divide" in health, ensuring that relevant information—and the technologies to deliver it—are widely available and effectively used by health personnel including professionals, researchers and scientists, and policy makers. Launched by the Secretary General of the United Nations in September 2000 and led by the World Health Organization, HINARI has brought together public and private partners under the principle of ensuring equitable access to health information. HINARI provides access to more than 2,000 scientific publications on public health to institutes within 113 eligible countries.

The Ptolemy Project enables surgeons in several Africa countries access to one of the world's largest libraries of surgery journals by designating medical researchers and clinicians in these countries as research affiliates of the University of Toronto. The University of Toronto Library is the third largest research library in North America and, as affiliates, participants are provided with full access to the Library's comprehensive collection of electronic resources.

What is becoming increasingly clear from these initiatives is the very strong need for both technology and training to properly exploit and utilize international information resources.

Of the more familiar open archive initiatives in the North many are either not directly relevant in the developing world (the high-energy physics arXiv) or where they are relevant, lack of awareness means access and use is low.

We must also be aware of evidence, as Richard Horton suggested, that the influx of information from the North to the South is in danger of damaging indigenous initiatives. Authors in the South are looking for recognition and reach as much as authors in the North and they may be tempted to publish in a journal with higher international "standing" than in a locally published journal, however relevant. This practice becomes a self-fulfilling prophecy where the lack of regard for local journals means they struggle with quality and in some cases with publishing at all.

South-to-South information flows are constrained by a lack of funds. High costs hinder awareness, use, and development of information resources. There is also a perception problem that perhaps the most relevant information arena is the North. The Association of African Universities' DATAD project, which publishes doctoral and master's theses from African institutes for African institutes, may prove to be an important example working against the trend here.

Can open access bridge any of these gaps? Let us look at the strengths, weaknesses, opportunities, and threats of open access in the context of developing countries. Open access has many strengths. First, it confers visibility. A recent study in *Nature* suggested that self-archived papers are cited "an average of 336 percent more citations to online articles compared to offline articles published in the same venue."[2] That kind of visibility could improve awareness of and access to resources. In terms of affordability and sustainability open-access software (such as those provided by the Eprints, DSpace, and CERN initiatives) and information are free. The distributed server technology encouraged by the Open Archive Initiative obviates the need for expensive local infrastructure. Another strength of open access is that the software is easy to obtain and straightforward to utilize. If archives are built with Open Archives Initiative compliance, it increases their potential for retrieval and for visibility as well as interoperability.

There are also some weaknesses associated with open access, most of which will likely be solved in the medium to long term. Open access initiatives and open archiving are technology-based concepts. In countries where there is poor ICT infrastructure and a relative lack of open access skills, this is a problem. Another weakness is that archives are presently rather North biased. The vast majority of archives in open access schemes are based in the North. Holding data in the North will potentially perpetuate access and use problems in the South. For example, the Botswana HIV Institution, which partners with Harvard, is concerned that because the data are principally housed in Harvard the Institution lacks control over those data and cannot guarantee their own access in the future. In addition, the content at some open-access sites is not peer reviewed, which presents a potential danger, for example, in medical disciplines. There is also the issue of quality control, involving peer review, the consistency of peer review, and version management.

[2]S. Lawrence. 2001. "Free Online Availability Substantially Increases a Paper's Impact," *Nature* 411 (6837):521. http://www.neci.nec.com/~lawrence/papers/online-nature01/.

Open access also presents many opportunities. Open access enables researchers to communicate their results widely, quickly, and cheaply. That is very important for researchers in developing countries. It also assists in connecting global research communities. Researchers and information users in the South often complain of feeling isolated from global information exchange and the wider scientific community. Open access also provides the potential to foster partnerships and strengthen scientific cooperation South to South and South to North.

Open access undoubtedly improves communication. This can assist in changing the perception in the North that only Northern research is relevant by exposing Northern researchers to the valuable information produced in the South. It also challenges "West is best" attitudes that sometimes pervade in the South. Open access speeds up what Jan Veltrop at Biomed Central calls the "minutes of science," the idea that with quicker and wider availability, there will be quicker and wider impact. This confers the potential to speed up the rate of progress, which could potentially benefit developing communities.

In addition to opportunities open access also poses some threats, including the threat of too much information. It may sound contradictory in countries where there has not been enough information, but researchers may find that they do not know which information resources are available in the online medium. When they do find it, they may not know whether it is worth having. Very little guidance about quality and reliability is typically available with online resources and open archives will have to be rigorous and dependable if they are to be worthwhile.

There is also a threat of researchers and information professionals in the South being, or at least feeling, left behind. The technology, the know-how, and the innovation are certainly being driven in the North, at the moment, potentially to the exclusion of the South. For example, in an informal survey of library professionals in about 13 African countries not one had heard of the Open Archives Initiative.

There are also concerns with archiving and perpetual access. The Botswana Harvard example illustrates these threats. Who is going to host the data and therefore who controls access in the long run? Who is going to pay for it and for how long? Developing countries are tired of donation programs that stop and start at whim, and open access could potentially have that problem. Copyright transfer agreements also pose a great challenge. While some publishers allow authors to self-archive under current copyright transfer agreements, in the future they may limit the rights of authors to post.

OPEN-ACCESS INITIATIVES IN DEVELOPING COUNTRIES

There are many initiatives that address open access in the context of developing countries. These include the Bioline[3] and the *British Medical Journal* initiatives, which provide 29 journals free to the world's 100 poorest countries. Another example is Extramed, which is based in the United Kingdom and takes its name from over 300 biomedical journals that are not indexed in Medline and therefore are invisible, even if they contain good and important research. When they are not seen, they are not cited. If they are not cited, their impact is reduced. It is a familiar story. Extramed publishes the full text of those journals on a CD-ROM and makes it available free to the South and at a reasonable cost to the North. Other projects include the *Electronic Journal of Biotechnology* in Valparaiso, and the INDMED database project in India. The Ptolemy Project[4] is concerned with measurability and the impact of their project. The Public Knowledge Project provides open access software for archive building in developing countries.[5] The SciELO project, based in Latin America, is a collection of online journals from Brazil, Chile, and Cuba that publishes in Portuguese and Spanish and provides abstracts in English and French. Journals from Costa Rica and Venezuela will be added to this project. The TerraLib project is an example of open-source software from Brazil.

INASP has compiled a directory of open access projects, key organizations, and related sites.[6] Currently, there are 65 projects—individual journals, subject repositories, journal collections, and databases—listed in that

[3]See Chapter 14 of these *Proceedings*, "Bioline International and the *Journal of Postgraduate Medicine*: A Collaborative Model of Open Access Publishing," by D. K. Sahu and Leslie Chan.

[4]See Chapter 12 of these *Proceedings*, "The Ptolemy Project: Delivering Electronic Health Information in East Africa," by Massey Beveridge.

[5]See Chapter 32 of these *Proceedings*, "Public Knowledge Project Open Journal System," by Florence Muinde.

[6]See http://www.inasp.org/peri/free.

BOX 29.1 Components of PERI

Component 1: Deliver Information
- Facilitate access to international information and knowledge

Component 2: Disseminate National and Regional Research
- Support access to national and regional publishing, strengthen and develop access to national and regional journals as a medium for the dissemination of local information and knowledge

Component 3: Enhance ICT Skills
- Enhance awareness and provide training in the use and/or evaluation of ICTs

Component 4: Strengthen Local Publishing
- Enhance skills in the preparation, production, and management of scholarly publications

directory, which includes a short description plus a Web link to the projects. This directory is constantly updated and is available on the INASP Web site.

INASP PROGRAMME FOR THE ENHANCEMENT OF RESEARCH INFORMATION

The Programme for the Enhancement of Research Information (PERI) is a four-component program of INASP (see Box 29.1).[7] At the moment Component 1 of the program offers Southern institutes access to high-quality, peer-reviewed information from publishers such as Oxford University Press, Gale, Emerald, Springer, Blackwell, and EBSCO that would cost a single institute in the North up to $2.4 million. That information is made available on a countrywide basis. For a single country in the South the costs are between $22,000 and $65,000. This component of PERI enhances North-South information flow.

It is clearly also important to support indigenous publishing and the South-to-North and South-to-South information flows. INASP started Component 2, the Journals OnLine Programme, to increase awareness of and access to research results nationally and internationally to stimulate author submissions, and to support subscriptions. The African implementation of this component, African Journals OnLine (AJOL), offers over 170 African journals online with tables of contents and abstracts. There are links to the electronic full text whenever it is available. When electronic delivery does not exist, INASP facilitates document delivery by fax or e-mail, with fair recompense to the publishers in Africa. AJOL also supports publications in local languages, such as Portuguese journals from Mozambique and French-language journals from the west coast of Africa.

If requests come from the South, they are fully subsidized by INASP, so that there is no cost to the user. If they come from the North, there is a U.S. $10 charge per article. The project is about visibility, about being seen, cited, used, and valued. The pages viewed per month has grown from 6,000 at the launch to 81,000 at the end of 2002. Since the project is trying to address North-South, South-South, and South-North communication, it tracks registrants by country or region. There are over 1,300 registrants based in Africa, over 1,000 in the United States, over 1,000 in Europe, and a good number from Canada, Australia, and Asia. There are approximately 4,800 registered users, and that number is growing.

INASP wants to enhance capacity building with developing countries by enhancing the skills and attitudes in the preparation and production of journals and in the medium term by moving this project to local management.

[7]See http://www.inasp.info/peri/index.html.

The project includes AJOL, which is the most established, and a pilot phase in the Caribbean, called Carindex. INASP is currently conducting feasibility studies in Nepal, Sri Lanka, and Bangladesh to establish online journal programs in each of those countries.

FUTURE DIRECTIONS FOR OPEN ACCESS IN DEVELOPING COUNTRIES

There is a shocking lack of awareness in the North and in the South about the potential of open access. It is important to raise awareness of open access and the related initiatives through conferences, workshops, and training. We should lobby for improved infrastructure, not just connectivity cables, hardware, and software, although these things are extremely important. We should highlight information access to policy makers as a priority. INASP has had some success with this in its programs in Uganda and Kenya. For example, the University of Makarere in Kampala recently doubled its bandwidth and is paying for more cabling to accommodate the information that is being made available through INASP's PERI program.

Equally important is training for capacity building. INASP runs in-country training courses that cascade throughout the country, because they train the trainers. INASP is only one initiative; there are others. INASP is teaching general ICT and Web skills as well as electronic journals management skills to library staff and faculty. There is arguably a need for specific open-access training: building repositories and understanding the protocols, data exchange, and archiving and retrieval.

There should be focused funding for open access projects, so that the South is truly included in the process. If there is going to be an open access revolution, it should involve everyone. Funding for projects in the South and partnerships with the North would be very helpful.

30

Open-Source Geographic Information Systems Software: Myths and Realities

Gilberto Câmara[1] and Harlan Onsrud[2]
National Institute for Space Research, Brazil, and University of Maine, United States

INTRODUCTION

The development of open-source software has received substantial attention recently. Following the successful examples of projects such as Linux, Apache, and Perl there has been substantial interest by policy makers and researchers on the dynamics of the production of open-source software (Benkler, 2003). A topic of particular interest is the adoption of open-source software systems in developing nations, as a means of reducing licensing costs and of promoting indigenous technological development by having access to the source code of these systems. A recent survey on intellectual property rights and international development commissioned by the government of the United Kingdom underpins such policies with an explicit recommendation.

> Developing countries and their donor partners should review policies for procurement of computer software, with a view to ensuring that options for using low-cost and/or open-source software products are properly considered and their costs and benefits carefully evaluated. (Barton et al., 2002)

Many studies that discuss the development of open-source software portray an idealized view that considers such software to be a product of a committed group of individuals. These individuals would operate on a distributed network, where each programmer works on a small but meaningful module. The programmers are isolated, communicating by means of a central repository and mailing lists. The incentives to participate operate on an individual level (Weber, 2002). Some authors go as far as identifying in open-source software a new mode of organizational structure denoted by commons-based peer production (Benkler, 2003). Others claim that the globally distributed skill induced by open source will loosen the grip of the richest countries on innovation (Kogut and Metiu, 2001).

This paper analyzes in detail one segment of the open-source software market in an attempt to find out the true extent of such claims and to establish the basis for a realistic view of the open-source movement. We will focus on geoinformation technology, which includes geographical information systems (GIS), location-based services, and remotely sensed image processing. We have chosen the geoinformation market for two main reasons. First, it is a key technology for developing nations, given its vast range of applications in areas such as environmental protec-

[1]Director for Earth Observation, National Institute for Space Research (INPE), Brazil. Web: http://www.dpi.inpe.br/Gilberto.
[2]National Center for Geographic Information and Analysis and Department of Spatial Information Science and Engineering, University of Maine, Orono, ME 04469-5711, USA, onsrud@spatial.maine.edu.

tion, urban management, agricultural production, deforestation mapping, public health assessment, crime fighting, and socioeconomic measurements. Secondly, the authors are experts on the area, with a substantial experience on geoinformation software development, and are in a qualified position to assess the different products.

We consider the following questions: (1) What are the conditions of open-source software development? (2) Who builds geoinformation open-source software products? (3) Is there a need for innovative open-source software applications in geoinformation applications? (4) How can developing countries obtain geoinformation open-source software to meet their national needs?

Our survey indicates that the view of open-source software as a product of a team of committed individuals is not realistic, at least for the geoinformation market. Most products are built either by a very small team of individuals or by corporations, and large collaborative networked teams are responsible for a small number of products. Most projects reverse-engineer existing designs or comply with standards, and few products are innovative. Therefore, there is much scope for new ideas, especially considering recent advances in geographical information science and spatial databases and the much-increased availability of Earth observation satellites. Given the constraints in open-source software production, such advances will not happen spontaneously and will require public intervention to fund innovation.

In order to support our claims we first examine the need for innovative geoinformation tools. We consider different models of open-source software production from an intellectual property viewpoint, and then review the process of open-source geoinformation software production. Lastly, we propose a model for open-source projects in the developing world based on networks of government-financed institutions.

THE NEED FOR INNOVATION ON GEOINFORMATION TECHNOLOGY

One of the motivations for our survey on open-source GIS software is to identify the extent of innovation in the community. There are three main drivers for innovation in geoinformation technology: (1) the evolution of database management systems to handle spatiotemporal data types; (2) the availability of a new generation of Earth observation satellites; and (3) the recent advances in geographical information science.

The complete integration of spatial data types in database management systems is bound to change completely the development of GIS technology, enabling a transition from the monolithic systems of today (that contain hundreds of functions) to a generation of spatial information appliances, small systems tailored to specific user needs (Egenhofer, 1999). Coupled with the data-handling capabilities of a new generation of database management systems, rapid application development environments will enable the construction of "vertically integrated" solutions, directly tailored to user needs. Therefore, an important challenge for the GIS community is finding ways to take advantage of the new generation of spatially enabled database systems to build "faster, cheaper, smaller" GIS technology.

A second reason for developing open-source spatial analysis tools is the need to resolve the "knowledge gap" in the process of deriving information from images and digital maps. This knowledge gap has arisen because our capacity to build sophisticated data-collecting instruments (such as remote-sensing satellites, digital cameras, and GPS) is not matched by our means of producing information from these data sources (MacDonald, 2002). To a significant extent we are failing to exploit the potential of the spatial data we collect. For example, there are very few techniques for image data mining in remote-sensing archives, and thus we are failing to use the information available in our large Earth observation data archives. Much of this knowledge gap has resulted from a substantial imbalance in public expenditure in geoinformation technology. Major Earth observation satellite programs such as ENVISAT and EOS have budgets in the billion-dollar range, where the vast majority of the money is spent on building and operating the satellites and sensors.

An additional challenge is how to incorporate recent advances from geographical information science into mainstream GIS. A number of important results have been produced in research areas such as spatiotemporal data models (Erwig et al., 1999), geographical ontologies (Fonseca et al., 2002), spatial statistics and spatial econometrics (Anselin et al., 1999), cellular automata (Batty, 2000), and environmental modeling (Burrough, 1998). These results have largely been outside of the reach of the user community because of a lack of widely available tools and systems that support them.

MODELS OF INFORMATION PRODUCTION IN OPEN-SOURCE SOFTWARE

From an intellectual property viewpoint we distinguish three models of information production for open-source software: (1) the postmature model; (2) the standards-led model; (3) the innovation-led model.

The postmature model exists in strongly consolidated markets. In many cases one proprietary product has a very large market share. As this product becomes popular its functionality and conceptual model becomes well established, and it becomes part of the public commons. Switching costs will prevent a new commercial product from capturing market share even if sold at lower prices. In this case there is a strong incentive for newcomers to license their products as open source. Many users will consider that the perceived benefits of open source will outweigh the cost of switching from the commercial product they might be using. One example is the Open Office productivity suite. Alternatively a private corporation may decide to license a product previously associated with private intellectual property rights as open-source software. Such is the case for the Mozilla browser.

The standards-led model exists when the establishment of standards consolidates a technology and allows compatible solutions from different producers to compete in the marketplace, thus opening an opportunity for open-source products. Newcomers can benefit from the substantial intellectual effort that goes into establishing a standard. An example is the SQL database standard, which has motivated products such as mySQL. Another example is the POSIX standard for operating system interfaces, which has reduced switching costs from other UNIX-based environments to Linux.

The innovation-led model results when universities, public institutions, and corporations produce work that has no direct equivalent in the commercial sector. As we shall see later, innovation is the product of the private sector, either directly (e.g., the Qt multi-platform interface system) or by a spin-off of a successful research project. As an example of the latter the University of California developed the Postgres database management system as a research project (Stonebraker and Rowe, 1986). After an unsuccessful commercialization attempt a private company took over the development of Postgres, added SQL support, named the resulting product PostgreSQL, and made it available as open source.

WHO BUILDS OPEN-SOURCE GIS SOFTWARE?

In order to conduct a more detailed analysis of the GIS open-source software developers, we conducted a survey of 70 GIS open-source projects, mainly using a listing provided by the freegis.org site,[3] a repository for open-source software. Based on size, geographical distribution, and affiliation we distinguished three categories of open-source software-development teams:

1. *Individual-led projects.* The project team consists of one to three individuals, usually from the same location and working in their spare time. The software products usually are small specialized applications that address specific requirements. In general the developer of the software is also its first user. Examples include the Vis5D visualization tool (Hibbard et al. 1994), the Gstat geostatistical package (Pebesma and Wesseling, 1998), and the shapelib library for reading ArcView® shapefiles.
2. *Collaborative networks.* The project core team consists of a team of 15 to 30 individuals geographically distributed. The developers usually have a separate job and do their work in their spare time, or in part-time allocated in agreement with their employer. Examples include the GRASS spatial analysis toolkit and the R collection of statistical functions.
3. *Corporation-based.* The project core team is part of an institution and is usually a group of three to eight programmers. There can be outside collaborators, but the main design decisions are made within the institution and in some cases should also address the commercial objectives of these corporations. Examples include the PostGIS extension to the PostgreSQL database management system, and the TerraVision systems for terrain visualization on the Internet.

[3]For additional information, see the FreeGIS project at http://www.freegis.org/.

We characterized each product according to its intellectual-property model and its development team. The results contradict the naïve view of open-source projects as a product of committed teams, based on peer pressure. More than half of the projects are led by individuals, and only four (6 percent) are based on a loose network of collaborators. The presence of corporation-based projects is very strong, with 41 percent of all cases examined. The results are further proof that all software, either open or closed source, is constrained by the essential properties of its development process: conceptual design, program granularity, cohesion of the programming team, and dissemination strategy.

The relatively small proportion of innovative projects (19 percent) indicates that the design of most open-source software products is based on the postmature and standards-led production models, where the main aim is not directly to produce innovation but to lower licensing costs and to break commercial monopolies. The strong presence of standards-led products is also a direct reflection of the influence of the OpenGIS consortium in the developer's community. This result further illustrates the notion that the hardest part of software development is the conceptual design of the intended product (Brooks, 1987). The two innovative projects developed by a networked team of programmers are GRASS and R. Both products have a simple and well-understood conceptual design, and their innovative contribution lies not in their design but in the analysis functions that scientists develop using these environments.

Out of the 29 corporation-based institutions involved in developing open-source GIS, 17 are private companies, 8 are government institutions, and only 4 are universities. This result indicates that the research community is usually not interested in a direct involvement in long-term, open-source projects. Maintaining and supporting an open-source software project requires considerable resources beyond the reach of most university groups. For a research prototype to evolve into an open-source product a team of developers must take over from the original research team and establish a support and maintenance infrastructure for the product.

Problem granularity is another important factor for open-source projects, and each type of software induces a different breakdown strategy. In most cases there is a strong limit on module size, which forces successful open-source products to be the products of small teams. The fact that GRASS consists of a set of independent executables is evidence that open-source development by distributed teams requires a software structure that can be broken into small, manageable parts.

Our survey of the open-source GIS projects also considered the maturity, support, and functionality of each product. We measured the maturity of a project by three factors: (1) the number of software releases; (2) the amount of changes in each release; and (3) the achievement of the project's stated goals. For assessment of support we investigated whether the project had an established maintenance team, and evaluated the mailing lists, bug indicators, and improvement requests. Evaluation of the concept of functionality considered the number of modules and the difficulty of the algorithms involved. Each project was graded on a scale from 1 to 5, where 5 is best.

The results indicate a significant difference in all three aspects (maturity, support, and functionality) between individual-led products and corporation-based ones. This indicates that the corporate environment is better suited for long-term software development than an individual's perspective. Individuals are constrained by their duties, which very rarely include a full-time support for open-source software development, whereas many corporations rely on earning indirect revenues (e.g., consultancy fees) from their open-source products. In many cases the corporation might be performing a public service or developing the product based on public funding. The results also indicate that the difference between a corporation and a collaborative network team is much smaller. This is consistent with the overall picture of the open-source world, that a committed team of individuals can produce results that are comparable (or better) than that produced by corporations.

USING AND PRODUCING OPEN-SOURCE SOFTWARE IN DEVELOPING NATIONS

The preceding sections have examined the nature of open-source software development and outlined the main characteristics of its production. We have argued that most mature and successful products require the establishment of organizational structures dedicated to their production. The consequences for developing nations are significant. Many developing nations are currently actively considering policies to support or enforce the adoption

of open-source software by public institutions (Dravis, 2002). The arguments in favor of adoption by public institutions include (Ghosh et al., 2002):

- *Lower cost.* Adoption of personal computers based on open-source software for public use can reduce initial entry cost by as much as 50 percent. Easier replication of solutions is also possible. Large-scale public projects can greatly benefit from having a prototype developed and tested that can then be replicated across the country with no additional software costs.
- *Independence from proprietary technology.* Many governments are increasingly concerned with overdependence in some important markets on a small number of vendors.
- *Security.* Governments and governmental agencies are becoming aware of the risks they are subject to when adopting proprietary software solutions in sensitive areas, such as e-government, e-procurement, elections, and public finance.
- *Availability of efficient and low-cost software.* The virtuous examples of some products (such as Linux and Apache) have encouraged statements about the widespread availability of open-source software for public use.
- *Ability to develop custom applications and to redistribute the improved products.* Given the open nature of open-source software, skilled local programmers could adapt the software to fit local needs and thus increase the efficiency of the services provided by the improved products.

The authors consider that there is enough empirical evidence to support the first three claims, but the issues regarding software availability and ease of customization are far more problematic and require a much closer examination. Most successful open-source software tools are infrastructure products, such as operating systems, programming languages, and Web servers. By contrast, the number of mature open-source, end-user applications is much smaller (Schmidt and Schnitzer, 2002). Operating systems, compilers, and Web servers respond to the needs of technically qualified information technology professionals, who can more easily adapt to the demands of products where support might be available only on the Internet and might require expertise in the English language.

There is a huge demand for end-user applications in developing nations, especially in the public sector. However, our survey indicates that corporations dominate open-source software development. These corporations will develop software based on their strategic interests, which are unlikely to include the full range of end-user applications needed by developing countries. Therefore, if governments in developing nations aim to profit from the potential benefits of open source, they must intervene and dedicate a substantial amount of public funds to support the establishment and long-term maintenance of open-source software projects.

The benefits of this strategy could be substantial. Consider, for example, the case of urban cadastral systems based on a spatial database for medium-size cities. The typical base cost of a commercial spatial database solution for one city is $100,000. If 10 cities were to adopt such a solution in a given year, there would be a saving of $1 million per year on licensing fees, which could finance local development and local adaptation.

There is also a substantial additional benefit of investing in qualified human resources. Government strategies for supporting indigenous open-source software development and adaptation would result in a learning-by-doing process. Such processes, as opposed to learning-by-using, are credited with fostering innovation in the developed world (Landes, 1999), and the same lessons could apply to those nations supporting emerging economies.

As an example of government-funded projects, a group of research and development institutions in Brazil is currently developing TerraLib,[4] an open-source GIS library that enables quick development of custom-built applications for spatial data analysis. As a research tool TerraLib aims to enable the development of GIS prototypes that would include recent advances in geoinformation science. On a practical side TerraLib enables quick development of custom-built applications using spatial databases. Projects such as TerraLib show that open-source GIS projects can make substantial contributions to the spatial information community by providing a platform for innovation and collaborative development (Câmara et al., 2000).

---

[4]See http://www.terralib.org.

CONCLUSION

This work examines the nature of open-source software development, by looking in detail at the application area of geoinformation technology. We surveyed 70 open-source GIS software projects and concluded that the Linux paradigm is the exception rather than the rule, and that corporations are the main developers of successful open-source products. Since networked teams develop only 6 percent of the all open-source GIS products, our result refutes the view that open-source software development defines a new mode of production. As established by extensive research, good software design and development are the products of qualified teams that operate at a high level of interaction. Developing software in a decentralized manner requires a modular design, which is difficult to achieve for most applications, since few software products can be broken in very small parts without a substantial increase in interaction costs.

The direct participation of universities in open-source software is limited due to the conflict between the generation of new research ideas and the need for long-term software maintenance and upgrades. As a result innovative projects account for less than 20 percent of the total and a large proportion of the projects (53 percent) simply aim to provide standardized components for spatial data processing. Individuals or small teams develop more than half of the products surveyed, and their best results are specialized applications aimed at conversion and visualization of data in established formats. Corporations account for 41 percent of all products and have a much better quality than individual-led software. This demonstrates that the impetus behind open-source software is not coming from altruistic individuals working in the midnight hour, but from professional programmers.

These results have important consequences for public policy guidance. First, good open-source software is the product of corporations, which will build them based on their strategic intents. Therefore, governments worldwide that try to benefit from the open-source software model by simply establishing legislation that mandates its use could be frustrated in their objectives, because of the lack of suitable public-sector applications. In order to create the software they need governments need to establish public-funded projects for open-source development and adaptation to local needs. Failure to understand the open-source development model will result in a lost opportunity for the developing world to reduce the current technological gap between the rich and poor nations.

REFERENCES

Anselin, L, P. Longley, M. Goodchild, D. Maguire, and D. Rhind. 1999. "Interactive Techniques and Exploratory Spatial Data Analysis," *Geographical Information Systems: Principles, Techniques, Management and Applications.* Geoinformation International, Cambridge.

Barton, J., D. Alexander, C. Correa, R. Mashelkar, G. Samuels, and S. Thomas. 2002. *Integrating Intellectual Property Rights and Development Policy,* UK Department for International Development, London.

Batty, M. 2000. "GeoComputation," in *GeoComputation Using Cellular Automata,* eds. S. Openshaw and R. J. Abrahart, Taylor & Francis, London, 95-126.

Benkler, Y. 2003. "Coase's Penguin, or, Linux and The Nature of the Firm," *Yale Law Journal* 112, winter 2002-2003.

Brooks, F. 1987. "No Silver Bullet: Essence and Accidents of Software Engineering," *IEEE Computer,* 20(4):10-19, April.

Burrough, P. 1998. "Dynamic Modelling and Geocomputation," *Geocomputation: A Primer,* eds. P. Longley, S. Brooks, R. McDonnell and B. Macmillan, John Wiley, New York.

Câmara, G., R. Souza, B. Pedrosa, L. Vinhas, A. Monteiro, J. Paiva, M. Carvalho, and M. Gattass. 2000. *TerraLib: Technology in Support of GIS Innovation,* II Workshop Brasileiro de Geoinformática, GeoInfo2000, Instituto Nacional de Pesquisas Espaciais, São Paulo.

Dravis, P. 2002. "A Survey on Open Source Software," The Dravis Group, San Francisco.

Egenhofer, M. 1999. *Spatial Information Appliances: A Next Generation of Geographic Information Systems,* First Brazilian Workshop on GeoInformatics, Campinas, Brazil, Instituto Nacional de Pesquisas Espaciais, Sao Paulo.

Erwig, M., R. H. Güting, M. Schneider, and M. Vazirgiannis. 1999. "Spatio-Temporal Data Types: An Approach to Modeling and Querying Moving Objects in Databases," *GeoInformatica* 3(3):269-296.

Fonseca, F., M. Egenhofer, P. Agouris, and G. Câmara. 2002. "Using Ontologies for Integrated Geographic Information Systems," *Transactions in GIS* 6(3):231-257.

Ghosh, R. A., B. Krieger, R. Glott, and G. Robles. 2002. *Open Source Software in the Public Sector: Policy within the European Union,* International Institute of Infonomics, University of Maastricht, Maastricht.

Hibbard, W., B. Paul, A. Battaiola, and D. Santek. 1994. "Interactive Visualization of Earth and Space Science Computations," *IEEE Computer* 27(7):65-72.

Kogut, B., and A. Metiu. 2001. *Distributed Knowledge and the Global Organization of Software Development,* Massachusetts Institute of Technology Press, Cambridge, MA.

Landes, D. S. 1999. *The Wealth and Poverty of Nations.* W. W. Norton & Co, New York.

MacDonald, J. 2002. *The Earth Observation Business and the Forces That Impact It,* Earth Observation Business Network 2002, MacDonald Dettwiler, Vancouver, CA.

Pebesma, E. J., and C. G. Wesseling. 1998. "Gstat: A Program for Geostatistical Modelling, Prediction and Simulation," *Computers & Geosciences* 24(1):17-31.

Schmidt, K. M., and M. Schnitzer. 2002. *Public Subsidies for Open Source? Some Economic Policy Issues of the Software Market,* Seminar for Economic Theory, Ludwig Maximilian University, Munich.

Stonebraker, M., and L. A. Rowe. 1986. "The Design of POSTGRES," in *Proceedings of ACM-SIGMOD International Conference on the Management of Data,* ACM Press, Washington, D.C., pp. 340-355.

Weber, S. 2000. "The Political Economy of Open Source Software," *Berkeley Roundtable on the International Economy,* University of California, Berkeley.

31

Open-Access Research and the Public Domain in South African Universities: The Public Knowledge Project's Open Journal Systems

Sal Muthayan
Doctoral Candidate, South Africa

INTRODUCTION

This paper, based on a broader empirical study I am conducting at three South African universities, examines how globalization and democratization have affected knowledge production at South African universities. It looks at whether increasing open access to academic research, which can be made readily available through new technologies, might enhance the research capacity of these universities and give impetus to the transformation project aimed at social justice and a new democratic order in South Africa. In this paper I consider the current state of research access at the three universities and whether open-access publishing systems, such as the Open Journal Systems developed by the Public Knowledge Project, are in a position to contribute to building research capacity at South African universities.

The questions posed in this paper are: (1) Given global economic trends and the low value of South African currency in the exchange market, what has been the access to scholarly resources and technology at South African universities over the past five years? (2) To what extent do the experiences of historically black universities (HBUs) differ from those of historically white universities (HWUs)? (3) What are the research access and capacity issues confronting researchers and librarians at South African universities? (4) To what extent can open access and the public domain of research be increased through the Open Journal Systems?

BACKGROUND AND CONTEXT

As a new African nation South Africa is faced with challenges that are both local and global. The country needs to build a democratic society from the devastation in the wake of apartheid while finding its niche in a globalized world that is knowledge-based (White Paper 1997, as cited in Kraak, 2001, p. 20). The new higher education policies and several scholarly analyses of this policy focus attention on the universities' agency in the transformation to a new democratic and equitable society nationally and in providing the country with skills, innovation, and knowledge to compete globally (Ekong and Cloete, 1997, p. 7; see also Currie and Subotzsky, 2000; Kraak, 2001; Soudien and Corneilse, 2000). Universally institutions of higher education have been seen as the main producers of knowledge and skills and as transmitters of culture (Sadlack and Altbach, 1997, p. 3; Neave, 2000, p. 1).

At the local level in South Africa the knowledge processes at the university can help to solve some of the many social problems facing the country. Universities can also assist in reconfiguring notions of culture, identity, and diversity in the postapartheid society (Cloete et al., 1997). In a knowledge-based global economy knowledge creation, innovation, and high skills formation at the university may help to position the country as a competitive global player. Although the new higher education policies emphasize the universities' dual role in these processes of democratization and globalization, little research has been conducted on whether these universities have the research capacity to generate this knowledge, innovation, and skills.

RESEARCH METHODS

The research methods consisted of an ethnographic multiple case study of research access and capacity conducted at three sites, namely, an HBU, a historically white Afrikaans university (HWU-A), and a historically white English university (HWU-E). Although the three cases were limited to one particular province in South Africa, their different social and historical contexts not only are similar to most other South African universities but also mirror the sociopolitical context of South Africa and the dilemmas it presents for the transformation of the society from an apartheid past to a democratic future. These cases have allowed me to examine any similarities and differences in research capacity between privileged HWUs and underprivileged HBUs in South Africa. Although the intention is not to generalize the findings of this research, as Miles and Huberman (1994, p. 29) evince, "Each setting has a few properties it shares with many others, and some properties it shares with some others, and some properties it shares with no others."

A Likert scale and short answers and in-depth interviews have been the main data gathering techniques in the survey. Thirty participants consisting of academics, postgraduate students, librarians, university administrators, and policy makers were surveyed and/or interviewed. Participants were drawn from faculties across the disciplines, from sciences to humanities and social sciences. The sample included racial, gender, and language diversity.

FINDINGS

Access to Scholarly Resources

Print Journals

Two of the three universities have experienced a sharp decline in library serial holdings (see Table 31.1). The HBU was most severely affected, with journal subscriptions across the disciplines having been cancelled due to financial constraints. Not a single new book had been acquired over the past five years. Although the academics and librarians at all three institutions place a premium on the value of research literature for teaching, supervising postgraduate students, conducting and publishing research, and providing counsel to public bodies, they have had to contend with this lack of access to scholarly resources. Interestingly, the HWU-E, an institution traditionally

TABLE 31.1 Serial Holdings

University	HBU	HWU-A	HWU-E
Print journal subscriptions			
2002	462 (45.6% decrease)	1057 (32.2% decrease)	1300
1997	850	1559	1300 (approx)[a]
Electronic database subscriptions 2002	14	13	21

[a]This university could not provide accurate figures of print subscriptions over the five-year period.

recognized for its research output, has managed to maintain its journal holdings. Participants at the HWU-E described their access to journals as adequate to good, at the HWU-A as adequate to poor, and at the HBU as poor. Invariably African journals were considered less important than international journals at the HWUs and equally important at the HBU. The discipline with the highest journal subscriptions was social sciences for the HBU, life and physical sciences for the HWU-A, and humanities for the HWU-E. Science and technology were among the disciplines with the lowest percentage of journal subscriptions. On average 80 percent of the holdings of these universities consisted of international (Western) journals, whereas only 20 percent comprised African journals, these being mainly South African journals.

One way that South African universities have overcome the limited access to scholarly resources is to form national and provincial consortia of university libraries. There were almost no endeavors to form regional African library consortia, a method used by other jurisdictions for securing greater access to journals (DFID, 1999).

Interlibrary Loans

Given the constrained access to scholarly resources, academics and postgraduate students had to depend heavily on interlibrary loan systems, which they believed was a further drain on their limited research funding. Articles retrieved internationally cost 110 rands while books cost 200 rands each. Postgraduate students working on their theses found this system "frustrating and time-consuming." One student explained the frustrations with such a system.

> Sometimes when I get that book, my study is not there. I have moved on and I am busy with other areas. Sometimes I feel I have wasted their [librarian's] time because I will refer maybe to a paragraph or two, but if I had made contact with that information from the beginning, I would have learned more. It posed [frustration] and it limits you and it makes you to make wrong choices because the tendency is that you are driven by the text (that is available) to make decisions. By the time you get, maybe things that are brilliant . . . you have closed your mind and made your choice and that is how you are going to do your research. So in a way I would say "yes, it does limit one."

This view shows that while interlibrary loan systems have worked well over the years, they do have shortcomings, which according to the student cited above, could have a significant effect on the research product.

Electronic Journals

Electronic access to journals has been well received by most participants. They cited the relative ease of wider access to scholarship and saving of time as the main advantages of electronic journals. In addition, several users may access an e-journal simultaneously, whereas print journals can only be consulted by one user at a time. For those who have access to the Internet at home, research may be conducted outside of library hours.

It is perhaps no surprise that the librarians of the universities with better resources—the HWUs—claimed that the most popular format for consulting research literature was online, whereas the librarian of the HBU posited that print form was the only format available to students. The HBU university's library had a total of 73 computers with Internet connections, yet only librarians used these computers (see Table 31.2). They conducted searches on behalf of academics and students, who had little direct access to them. The HWU-A's library had a total of 94 computers

TABLE 31.2 Library Computers with Internet Connections

University	HBU	HWU-A	HWU-E
Computers with Internet connections, 2002	73	94	102
Student use	0	8	49 (including public)
Faculty use	0	0	nil

with Internet connections but only 8 were available for use by the students and none by academics. The HWU-E had approximately 102 computers with Internet connections used by librarians and staff, 49 of these being available to the public, including students. At the HWUs the academics had their own computers with World Wide Web access. At the HBU, which is rurally located, the academics had only recently acquired computers with World Wide Web access, but the quality of the connections was still problematic and often resulted in slow and/ or aborted connections.

Concerns about Access to Scholarly Resources

Reasons for the declining access to scholarly resources went beyond the financial capacity of the individual universities. Librarians and academics at the HBU and the HWU-A complained that decision making about the allocation of resources seemed to indicate that management had placed a low priority on research output. These participants contended that a strong research culture, such as the one at the HWU-E, was not prevalent at their institutions. Policy makers and senior managers at the HWU-E placed a strong emphasis on research in their planning and budgeting. The dean of research, an active researcher, was also a member of senior management. In addition, the head librarian was included in budget discussions

Participants at the HWU-A attributed this lack of research culture to the history of these universities. The former government of South Africa had established these racially segregated universities to implement and consolidate its apartheid policies. By and large the HWU-Es produced skills for the mining and manufacturing industries (Bolsmann and Uys, 2001; Mabokela, 2000; Nordkvelle, 1990). HWU-As produced an Afrikaner elite to assume key positions in politics, government, and public administration, while HBUs were intended to legitimate the policy of separate development and to reproduce the subordinate social and economic position of black people (Mabokela, 2000; see also Subotsky, 2001). Furthermore, HBUs received an inequitable resource allocation from the apartheid state. Between 1989 and 1990 just before the end of apartheid, the 10 white universities spent more than 300 million rands on research while the six black universities spent a mere 24 million rands (Nordkvelle, 1990, p. 10). The new policies aimed at transforming higher education in South Africa have now emphasized the importance of research (White Paper, 1997) and individual institutions are beginning to institute significant incentives and rewards for research output by academics. At the HBU, for example, the university may receive 30,000 rands from the government for articles published in rated journals, with 12,000 rands of this going to authors (interview with executive dean, HBU, 2002).

Librarians and academics also pointed out that although electronic access opened new pathways and allowed for wider access to resources, it was not necessarily cheaper than subscriptions to print journals. The costs of electronic scholarly resources were still prohibitive due to unfavorable currency exchange rates and dwindling resources. In addition, researchers and librarians believed that existing facilities and resources were not being used optimally because many academics and postgraduate students did not have sufficient information technology literacy skills to conduct advance searches for print or electronic resources.

This, they claimed, was particularly acute among some black academics and the majority of black postgraduate students who had little to no access to libraries or technology outside of the university. Whereas white academics and students reported having access to online connections both at the university and at home, most black academics and students had no access to these facilities at home. In fact, some academics at HBUs were not assigned or allocated their own computer. By contrast, as one participant observed, many white university students had access to computers at the secondary school level and were therefore more empowered to conduct their research. Black academics, on the other hand, were sometimes reluctant to admit their lack of search skills to the mostly white librarians at the HWUs. A black academic, who felt embarrassed about admitting her lack of search skills to the librarians, shared the empowering effect electronic access has had on her, enabling her to conduct the searches in the privacy and comfort of her office, away from the intimidating gaze of the librarians.

Library orientation programs, consisting of hour-long sessions, were inadequate for equipping students with necessary information and skills to use the facilities efficiently. As one student pointed out,

> That is your only orientation to the library. . . . It is not an individual one hour slot, but it is the whole group. Then you are offered that if you want to come back you are welcome. Yes it is good to say that but as a student it hinders you. . . . [They may] show me this one journal . . . I want different options that say if you can't find it here you will find it there. It is like they are holding the information to themselves and they are using that information against you as a student. That is how you think because it threatens you and you as a graduate student don't want to look stupid.

In their evaluation of these sessions postgraduate students expressed the need to have such sessions during their first year at the university. Librarians were particularly concerned that they did not have sufficient time to devote to information literacy training because of their workload. Inevitably the libraries at all three universities were short-staffed. Vacancies could not be filled because of a lack of finances. Existing staff was expected to fulfill a number of functions. For example, at the HBU the periodicals librarian was also a subject librarian and a cataloguer. Some participants—librarians, academics, and students—were of the view that information literacy should be integrated into coursework, "just like the research methods course" said a postgraduate student.

Research and Publishing

Almost all academics claimed that they devoted a greater percentage of their time to teaching than research. The general lack of resources for recruiting more staff resulted in most universities' increasing the teaching loads of academics. On examining the allocation of time spent on research, teaching, and administration and policy, the lowest ratio for research was 10:50:40, whereas the highest ratio was 60:20:20. On average, academics devoted only 20 percent to 30 percent of their time to research. When asked about their visions for the future, almost all academics expressed a strong desire to have more time for conducting and publishing research.

Aside from this lack of time for research, new academics, in particular black academics, and students asserted that they received very little support on how to go about publishing. One student was told by her supervisor that certain journals only accepted articles from their members. This would mean that black researchers might find it difficult to publish in a country like South Africa, where previous patterns of privilege still exist. A student mentioned the need to understand the conventions of publishing and the hindrance these conventions posed for novice researchers. Older academics expressed a preference to publish in international journals mainly, while newer academics, both black and white, claimed it was important to publish in South Africa or Africa, where their research might hold more relevance and thus serve a social purpose. On the other hand, participants also pointed out that locally published journals are not as highly rated as international journals and hence there is little incentive for publishing in them. One participant was concerned about the hegemony of the West in publishing and dissemination of knowledge, pointing out the anomaly that some African journals are published in the developed world and then sold to African institutions at exorbitant prices.

Open Access

All participants in the study expressed positive views about greater open access to research resources, believing that it would provide greater access to scholarly information, enable researchers to conduct searches and retrieve information with ease, reduce costs, and save time (see Table 31.3). The librarians in particular were positive about open access. Rather than consider it a threat to their jobs, they believed that their role in an open-access world would be to act as managers and facilitators of information.

The participants expressed a number of concerns regarding open access. One participant pointed out that open-access publishing would not offer the kind of financial incentive institutions or individuals receive from the government for articles published in rated journals. Hence, he could not foresee South African researchers publishing in open-access journals unless they were rated. But the main concern expressed by several participants related to the abundance of information available through an open-access system. Librarians and academics were concerned that the students may find the information overload overwhelming. Given the students' relative lack of information literacy they may not be able to distinguish quality research from the range of research materials and information available through open access. Academics were of the view that open-access scholarship must be

TABLE 31.3 Advantages and Concerns Expressed about Open Access

Advantages	Concerns
Greater access to journals and research information.	Quality and standards may drop.
More sources of information.	Peer-review systems are essential for online publishing.
Access to most recently published articles.	Issues relating to copyright and intellectual property rights may be problematic.
Home access to scholarly information.	Plagiarism.
Enables students to search and retrieve information by themselves without the help of librarians.	Open access may reduce credibility of journals.
Provides students with more choices.	Some scholars feel it is "beneath" them to publish online.
Inspires users to conduct more research.	Can only benefit a few with access to technology. Access to technology is still limited in South Africa, especially at HBUs.
Time and energy saving.	Speed of access still problematic where technology is inadequate.
Timely access as opposed to time-consuming interlibrary loans.	Lack of information literacy skills. Low capacity to use technology.
High cost of subscriptions will be a thing of the past.	Management and security of technological facilities and equipment is a problem.
Will help with problem of declining library subsidies.	Information overload. Need information management training on how to critically evaluate information.
Reduce the costs of publishing.	Universities receive funds for articles published in rated journals. Open access journals would have to be rated.
Reduces need to use personal funds for articles.	

subjected to the peer-review system and that students must be coached on how to become discerning users of open-access scholarly resources.

Although open access can be an "open door to learning," as one student put it, it could also result in researchers expending valuable time on irrelevant information, especially if they do not have good information literacy skills. Other concerns related to plagiarism, copyright, and intellectual property rights. Although many academics were very concerned that open access would lead to greater incidences of plagiarism among students, a few pointed out that the Internet has made it easier to monitor and detect plagiarism.

A further concern was that open access would increase the digital divide in South Africa. The following observation made by a study participant may pertain to the global realm as well: "People that have had access before, like the white people of this country, will still be more advantaged with technology because they have computer access in their homes and the very people that were discriminated are left behind now." She expressed a sense of hopelessness in all of this: "I think it is more threatening to them. People give up. . . . They feel more disadvantaged now." But she admits that we cannot ignore technology and that this sense of hopelessness may be limited to the older generation:

> The solution is to [begin with] the younger generation, to have more computers in schools. . . . I have learned with children that technology is not something that is threatening. It is a challenge to them, but to us who are grown-ups it becomes a threat. That is why I think the earlier people are exposed, the more advantages they will see in technology because technology is beautiful!

Although this statement ends on a positive note, the concern that open access would only benefit a small privileged group, rendering those without technological capacity to the periphery, presents a real challenge to proponents of open access. It is a challenge to which careful consideration should be given, even as we push the frontiers of open access.

Public Domain of Academic Research

The notion of the public domain of academic research was relatively new for some participants and they had some difficulty with its credibility. Others, however, believed that universities had a social responsibility and that the research generated by universities should have a public value. Several participants were not convinced that research could become readily available to the wider public—given their relative lack of access to technology—but they claimed that it was essential to at least make research available to practitioners and policy makers. As one participant noted,

> There are people out there, not all of them are interested in being researchers, but they are interested in being competent practitioners. If they could have access to the research that is being done that would give more solutions to the problems that they are encountering out there. . . . That would be a useful system.

Participants also pointed out that there were examples of community centers equipped with computers and Internet access throughout the country, rendering the notion of community access to knowledge less remote. However, the quality of connections and the incidence of power failures in remote areas posed a problem for electronic access to public knowledge.

Some participants were of the view that academic researchers should play a bigger role in shaping public policy, locally, nationally, and regionally. They contended that regional African organizations such as the Southern African Development Community and the New Partnership for Africa's Development could benefit from research that focused on regional growth, sustainable development, peace, and security. Open access would also ensure that the deliberations of these organizations at various conferences and forums, could be made available to the public immediately.

Over recent years the conditions attached to research funding stipulate that South African researchers work closely with a range of stakeholders, including local communities. A botanist pointed out that she was working closely with rural communities and the government in water and estuarine studies; a legal researcher explained that her research dealt with customary laws and the rights of African women and that this information needed to be made available to these women. Although these communities do not have electronic resources, the researchers explained that they popularized this research through community talks, radio, popular magazines, and pamphlets. Some academics worry that applied research of this nature will erode the base of pure research. Others contend that it is possible to conduct socially relevant research while maintaining the foundations of basic research. Researchers in the medical and pharmaceutical sciences were less hopeful about making their research open to the public because of intellectual property rights.

Librarians noted the increasing use of their facilities by the public. Corporate and professional bodies and individual lawyers, teachers, social workers, business people, and other interested persons were already using academic library facilities. Yet others expressed concern about Western domination of knowledge and definitions of knowledge and public knowledge. One participant, for example, raised the question about the imperialism of knowledge and knowledge ownership: "Who generates the knowledge? Who sets the rules and determines what is scientific?"

DISCUSSION

Given the high expectations South Africa has of its higher education sector in contributing to knowledge production, innovation, and skills development as noted above, the universities in this study do not have the necessary research access and capacity to fulfill their roles as agents of transformation. As has been the case for most of Africa and the developing world, there is a lack of adequate resources for producing knowledge, making innovative interventions, and developing a highly skilled workforce.

Access

Access to the latest international research through updated serial holdings and other research literature is key to producing cutting-edge research and finding innovative solutions to the range of social and development problems facing South Africa. Teaching and research are integrally related. A university cannot produce highly skilled professionals when academics and students do not have access to the latest research in their respective disciplines.

The findings in this study show that journal holdings have declined sharply at two of these universities, almost 50 percent at the HBU, where not a single new book has been acquired in five years. Journals in discipline areas that are crucial for innovation and development, science, and technology have suffered the most. The costs of electronic journal subscriptions have also been prohibitive. The interlibrary loan system, although an essential resource facility in higher education over the years, is not without its problems. The process can be slow and this might impact negatively on the quality of research produced, as noted in the findings above.

Despite the limited availability of resources, access to technology such as computers with Internet access does not appear to be a big problem. As can be seen from Table 31.2, all libraries had between 74 and 102 computers with Internet connections. Poor connections at the HBU recently have been resolved. Academics at the three universities have their own computers in their offices and some even have home access. The problem with the library computers is that students should learn how to use them independently of the librarians. It is no surprise that at the HWU-E, the university with the highest research output among the three universities, 49 of the 102 computers available were being used not only by students but the public as well. Perhaps the remaining two universities should also make such facilities more readily available to users so that they may develop the skills to use them independently.

Another consequence of inadequate resources has been understaffing, of both librarians and academics. Librarians cannot find the time to train academics and students on how to use existing scholarly resources and facilities optimally. This creates a double jeopardy; not only are the libraries limited in scholarly resources but existing materials are also not being used efficiently and optimally. As some participants wisely observed, information literacy programs should be integrated with the coursework, but this will require additional human resources to conduct this training. The historical background of these universities and the low prioritization that management and policy makers have given to research at two of these universities have affected their research capacity negatively. It seems that when there were cuts to spending, the libraries at all three institutions were always targeted (interviews with librarians and policy makers). The established research culture at the HWU-E may largely be attributed to the emphasis senior management and policy makers place on research activities.

For academics teaching has been prioritized, whereas the time afforded to research is viewed as a luxury or privilege for only a few. As noted the quality of teaching depends on the research being generated and vice versa. Heavy teaching loads mean that these institutions are generating little research. Since most academics expressed a strong desire to have more time for research and publishing, one wonders how this constraint may affect their sense of job satisfaction, self-worth, and their identities as researchers. Already South Africa, like many African countries, is experiencing a "brain drain." Over the five-month initial phase of this research two academic participants from different institutions, both in computer and information sciences, emigrated. In an earlier interview one, himself a dean, expressed much dissatisfaction with what he and other academics perceived as management's disinterest in research. The consequence of the "brain drain" is that it further erodes the research and skills base of a country like South Africa.

Research and Publishing

Publishing presented a dilemma for some South African researchers. Limited access to the latest research developments can seriously impinge on researchers' capacity to produce and publish research, let alone cutting-edge research (see also Altbach, 1987; Canagarajah, 1996). As Altbach contends, scholarly journals are a key element in the knowledge distribution network and are even more important than books (p. 72). Some participants believed that their research was more relevant to the South African or African context, for example, in estuarine

studies and the African woman's right of succession in customary law. Yet, there were higher incentives, such as status, recognition, and rewards, for publishing in international journals. On the other hand, African journals were not highly rated nor were they well publicized and some participants did not even know of their existence. This trend is common in other African countries as well (see DFID, 1999, p. 7). In general, African journals had not been subscribed to and were thus not available at the library. Given the limited resources librarians prioritized subscribing to international and South African journals over African journals. As noted above, many African journals are published in the West and sold back to African universities at high cost. Participants referred to this as the "imperialization of knowledge."

A few developed nations dominate the production and distribution of knowledge by controlling the publishing houses and the production of scholarly journals that the rest of the world consumes: 34 industrialized countries with only 30 percent of the world's population produce 81 percent of the world's book titles (Altbach, 1987, p. 18). Although these figures are dated, scholars seem to concur that the knowledge gap has increased and will continue to do so (Altbach, 1998; Gibbons et al., 1994; Willinsky, 2000). According to Altbach, 62 percent of social science periodicals and virtually all "prestigious scientific journals" are published in the West (p. 28). In addition, the spate of mergers and acquisitions in journal publishing over the last few decades appears to have set off spiraling price increases that are undermining the circulation of knowledge. These increases can be traced to a growing corporate concentration in scholarly publishing, especially in the sciences, which has resulted in three Western companies, Elsevier, Springer, and Taylor and Francis, controlling 60 percent of the journals in the leading citation index, ISI Web of Science (Merger Mania, 2003).

Hence, these Western countries define research paradigms and the focuses of the field, rendering the rest of the world peripheral in determining the research agenda (Altbach, 1987, p. 17; 1997, p. 16). As has been shown above, there has been little exploration of regional networks to overcome such barriers to publishing. Nor does foreign aid help (Altbach, 1987, pp. 17-27; Day, 2002, p. 3). Instead the university libraries in this study received large donations of not only irrelevant books but also hundreds of copies of the same book. In addition, scholars like Altbach have shown that the textbook publishing programs of the World Bank have actually weakened the indigenous private firms (1987, p. 24; 1996, p. 7). According to Altbach, neo-colonialism is maintained through foreign aid programs and loan policies and is a factor that must be considered in any analysis of publishing in the Third World (1987, p. 33).

Aside from a lack of time for research and publishing, black academics and students noted the lack of support and access to publishing. Publishing rules and conventions often inhibited them from publishing. Scholar Canagarajah (1996) refers to these conventions as the "'nondiscursive' requirements" of academic publishing houses in the West, a "hidden publishing" agenda that makes it virtually impossible for researchers from the Third World to publish successfully in the industrialized world and leads to the exclusion and marginalization of peripheral (Third World) research (p. 1). These requirements include format of copy text, bibliographical conventions, weight and quality of paper, number of copies required, postage, procedures for revision, procedures for interaction between author and board, and deadlines (p. 2). These conventions speak directly to scholars in the West alone, ignoring not only the context of peripheral writers (e.g., a lack of access to computers, photocopiers, fax machines, and telephones; electricity; copy paper according to specifications; funds for postage of bulky copies, especially to referees in the West; access to reference conventions; access to journals as guides and diskettes), but also time and space factors, which global technology has not yet compacted in most of the world. In setting deadlines of three to four days editors reveal that they have little concept of the distance between Sri Lanka, for example, and the West, or the unreliability of international mailing systems (see also Day, 2002).

Based on his experiences and that of fellow scholars at the University of Jaffna, Canagarajah (1996) shows that these conventions preclude peripheral scholars from publishing. More perniciously, the apparent lack of attention to these requirements on the part of peripheral scholars may result in them being labeled unscholarly, unprofessional, or downright incompetent, despite the substantive value of their research. He notes that scholars in Asia, Latin America, and Africa have similar experiences (p. 9; see also Muchiri, et al., 1995). Canagarajah contends that "these publishing conventions are deeply implicated in the politics of knowledge production and the hegemony of intellectual property of the developed nations" (p. 3). Drawing on Foucault, he shows that these rules of publishing serve to legitimate particular conventions and exclude others.

A more reciprocal flow of knowledge and publishing would not only benefit the periphery. Canagarajah argues that if all knowledge is situated and personal then periphery perspectives, which are often critical of center research, may enrich and expand the narrow knowledge base of the center (p. 21). Altbach (1996) recommends a better balance of the research agenda between researchers and users, the strengthening of regional and international networks for sharing of research, and the inclusion of peripheral research communities in the international mainstream (p. 20).

Open Access

This discussion focuses on participants' orientation toward open access and the public domain of academic research. As previously noted, the participants welcomed open access and the possibility of making academic research more publicly available. Their concerns about open access centered on the quality of the research and the need for strong peer-review systems, information literacy, and management systems to deal with the information overload, plagiarism, and inequitable access to technology that might lead to the exacerbation of the digital divide.

The participants expressed enthusiasm at the prospect that academic research could be made available to the public at large. Although they were concerned that rural communities would not have the technology to access such information, they agreed that it was worthwhile to make this information available to practitioners and policy makers, whose work impacts directly on the people "out there." The new higher education policies emphasize the social value of academic research and support applied research conducted in collaboration with public stakeholders (see Gibbons et al., 1994). Hence, the expansion of the public domain of research in South Africa may be well received when viewed as a contribution toward the democratization process.

The discussion above notes the constraints on current South African research capacity, leading to the following crucial questions for developing research capacity not only in South Africa or Africa but also in other parts of the developing world. What measures do we have that are readily available to deal with the factors constraining research access and capacity in South Africa, Africa, and the developing world? How can we overcome resource constraints and increase access to journals so necessary for the production, publishing, and distribution of knowledge? How can we begin to establish journals with locally relevant content and whose agendas are determined by periphery researchers and editors?

Open-access systems such as the Open Journal Systems of the Public Knowledge Project may be just one example of how open access may help to address the constraints the academics, graduate students, and librarians in my study have voiced. The following section discusses the possibilities of the Open Journal Systems in building research capacity in developing world contexts, such as in South Africa.

THE OPEN JOURNAL SYSTEMS[1]

The Open Journal Systems of the Public Knowledge Project,[2] the University of British Columbia's federally funded research initiative to improve the scholarly and public quality of academic publishing on a global basis, was launched in November 2002. The Open Journal Systems is an online journal management and publishing system that enables editors to manage, publish, and index peer-reviewed journals over the Internet on an open-access or free-to-read basis. It can be installed on Web servers anywhere and requires few if any technical skills from editors. It has tremendous potential to make journals easier, more efficient, and cheaper to run.

The Open Journal Systems open-source software can be downloaded free of charge. The intention is to enable journals and scholarly societies to consider publishing in an open-access or free-to-read basis, which has been shown to increase readership dramatically.

Open Journal Systems is currently under consideration in Canada, Turkey, Kenya, Rwanda, India, Australia, and the United States. In January 2003 it was listed as a "landmark event" in the timeline of the Free Online

[1]Additional information about the Open Journal Systems can be found in Chapter 32 of these *Proceedings*, "The Public Knowledge Project's Open Journal System," by Florence Muinde.

[2]See http://www.pkp.ubc.ca. A demonstration journal and further information about Open Journal Systems may be found at http://pkp.ubc.ca/ojs.

Scholarship movement by Peter Suber. The Open Journal Systems has the option of being able to publish by issue, volume, and year, by volume and year, and by year alone. It can be refined so that each article comes out by number and exact date within a given volume and year. As one interested person from Turkey observed, "I did not expect such a comprehensive programme for free!"

Possibilities for Building Research Capacity

As noted, resource constraints severely hamper the capacity of South African universities to produce and publish research. An online journal system such as Open Journal Systems can make journals readily available to academics, postgraduate students, and librarians at almost no cost. Academics and graduate students can have easy access to the latest research for both teaching and research purposes, including journals to which libraries currently have very limited access, namely, science and technology journals. The HBU, with its severely constrained coffers, would especially benefit from such a system.

Libraries could then use the savings from journal subscriptions, both print and electronic, to recruit staff to deliver information literacy courses, which would help students to use technology more efficiently, optimally, and independently of the librarians. Hence the HBU and HWU-A will be more confident about allowing their students to use the existing library computers. The course will also empower students to cope with the information overload, enabling them to become critical and discriminating users of online open access. The negative impact the slow processes of interlibrary systems may have on research will be reduced since the researcher will not have to depend on this system alone.

Although resources will be needed to recruit staff to lessen teaching loads, open-access publishing would cut down on the time spent accessing print and interlibrary loan materials, especially materials that may not be available through the libraries' e-journal subscriptions. The time saved can be allocated to conducting more research.

Open Journal Systems perhaps has the most to offer in the area of local or, as Canagarajah puts it, "peripheral" publishing. As we have seen in the discussion on publishing, not only do peripheral scholars have limited access to the latest research but the publishing conventions also inhibit them from publishing successfully. Peripheral scholars have no voice in defining research paradigms, focuses for the field, or what constitutes relevant research within a particular context or environment. Open Journal Systems has the potential to make the latest research readily available to all researchers at no cost through a fully indexed system. Peripheral editors may determine focuses and content that is context relevant. The conventions for publishing are easy to follow. This means greater freedom for peripheral scholars who have felt constrained by the "nondiscursive requirements" of publishing. Although some conventions still should be followed, other cumbersome requirements such as quality paper, copies, postage, communication with editors, and unrealistic deadlines for the resubmission of articles can all now be avoided through open online publishing using the Open Journal Systems. This system may be suitable not only for South Africa and the African continent but for countries in Asia and Latin America as well, where scholars have had similar constraints.

More importantly, publishing through an open system like the Open Journal Systems would perhaps enable a more reciprocal flow of knowledge between the center and the periphery, allowing peripheral perspectives, especially when considering developing world issues, to expand and enrich the narrow knowledge base of the center. This should not be done merely with the hope of including peripheral research in the mainstream but rather to revisit the notions of knowledge and knowledge ownership to confront what one participant referred to as "knowledge imperialism."

Concerns about open access should be addressed. There should be little concern about the quality of the research published through Open Journal Systems because the journals will be peer reviewed. Also, free access through the Open Journal Systems can potentially expand the public domain of research. For those who have access to technology, for example, practitioners and policy makers, Open Journal Systems can be a readily available source of scholarly information on which competent practice and good policy making may be based. The lack of access to computers with Internet connections need not be a major shortcoming at this point because we are seeking to increase access for those who have a growing ability to tap into technology but cannot afford the prohibitive costs associated with current print and e-journal access.

REFERENCES

Altbach, P.G. 1987. *The Knowledge Context: Comparative Perspectives on the Distribution of Knowledge,* State University of New York Press, Albany.

Altbach, P.G. (ed.). 1996. *The Challenge of the Market: Privatization and Publishing in Africa,* Bellagio Studies in Publishing, 7, Bellagio Publishing Network, Chesnut Hill, MA.

Altbach, P.G. 1998. *Comparative Higher Education: Knowledge, the University and Development,* Ablex Publishing Corporation, Greenwich, CT.

Bolsmann, C., and Uys T. 2001. "Pre-empting the Challenges of Transformation and Marketisation of Higher Education: A Case Study of the Rand Afrikaans University," *Society in Transition,* 32(2):173-185.

Canagarajah, A.S. 1996. "'Nondiscursive' Requirements in Academic Publishing," *Written Communication,* 13(4):435.

Cloete, N., Muller, J., Makgoba, W. M., and Ekong, D. (eds.). 1997. *Knowledge, Identity, and Curriculum Transformation in Africa,* Maskew Miller Longman, Cape Town.

Currie, J., and Subotsky, G. 2000. "Alternative Responses to Globalization," in Nelly P. Stromquist and K. Monkman (eds.), *Globalization and Education; Integration and Contestation Across Cultures,* Rowman & Littlefield Publishers, Oxford, pp. 123-147.

Day, C. 2002. "Globalizing the Exchange of Ideas," *The Chronicle of Higher Education,* 48(21).

DFID. 1999. "African Journals—an Evaluation of the Use of African-Published Journals in African Universities Evaluating Impact," *Education Research Paper No. 36,* see http://www.dfid.gov.uk./AboutDFID/Education/Research/Library/contents/c0173e/begin.htm

Ekong, D., and Cloete, N. 1997. "Curriculum Responses to a Changing National and Global Environment in an African Context," in N. Cloete, J. Muller, M. W. Makgoba, and D. Ekong (eds.). *Knowledge, Identity and Curriculum Transformation in Africa,* Maskew Miller Longman, Cape Town, pp. 3-16.

Gibbons, M., C. Limoges, H. Nowotny, S. Schartzman, Scott, and M. Trow. 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies,* Sage Publications, London.

Kraak, A. 2001. "Equity Development and New Knowledge Production: An Overview of the New Higher Education Policy Environment in Post-Apartheid South Africa," *Equity and Excellence in Education,* 34(3):15-25.

Mabokela, R.O. 2000. "We Cannot Find Qualified Blacks: Faculty Diversification Programmes at South African Universities," *Comparative Education,* 36(1), p. 95, 18p.

Merger Mania. 2003. *Scholarly Communication Reports*, 7(5):2-4.

Miles, M.B., and Huberman, A.M. 1994. *An Expanded Sourcebook: Qualitative Data Analysis,* Sage Publications, Thousand Oaks, London and New Delhi.

Muchiri, M.N., Mulamba, N.G., Myers, G.A., and Ndoloi, D. B. 1995. "Importing Composition: Teaching and Researching Academic Writing Beyond North America," *College Composition and Communication* 46(2):175-198.

Neave, G. 2000. "The Universities' Responsibilities to Society: International Perspectives," in Guy Neave (Series ed.), *Issues in Higher Education* series, International Association of Universities and Elsevier, Amsterdam.

Nordkvelle, Y. 1990. "The Academic Boycott of South Africa: Science and Social Practice," *Studies in Higher Education,* 15(3):253-276.

Pennycook, A. 1996. "Borrowing Others' Words: Text, Ownership, Memory and Plagiarism," *TESOL Quarterly,* 30(2):201-230.

Sadlak, J., and Altbach, P.G. (eds.). 1997. *Higher Education Research at the Turn of the New Century: Structures, Issues and Trends,* UNESCO, Paris.

Soudien, C., and Corneilse, C. 2000. "South African Higher Education in Transition: Global Discourses and National Priorities," in Nelly P. Stromquist and K. Monkman (eds.), *Globalization and Education; Integration and Contestation Across Cultures,* Rowman & Little field Publishers, Oxford, pp. 123-147.

Subotzky, G. 2001. "Addressing Equity and Excellence in Relation to Employment: What Prospects for Transformative Change in South Africa?" *Equity & Excellence in Education,* 34(3):57-69.

White Paper. 1997. As cited in Kraak, A. 2001. "Equity Development and New Knowledge Production: An Overview of the New Higher Education Policy Environment in Post-Apartheid South Africa," *Equity and Excellence in Education,* 34(3):15-25.

Willinsky, J. 1999. *Technologies of Knowing: A Proposal for Human Sciences,* Beacon Press, Boston, MA.

Willinsky, J. 2000. *If Only We Knew: Increasing the Public Value of Social Science Research,* Routledge, New York.

32

The Public Knowledge Project's Open Journal Systems

Florence Muinde
UNESCO Fellow, Kenya

INTRODUCTION

This presentation focuses on the activities of the Public Knowledge Project, particularly on developing country applications and issues. A federally funded research venture at the University of British Columbia, Canada,[1] the project's driving principle is that knowledge should be free, hence, their mission to promote a world where knowledge is free. Over the last four years the project has involved research on ways to improve the scholarly and public quality of academic research through the use of online technologies. We have been gathering data on the impact of open-access publishing in Cameroon and South Africa, as well as on policy makers and professionals in Canada.

In addition to political, economic, legal, and social analysis of open-access publishing issues, the Public Knowledge Project just completed an open-source software system, Open Journal Systems,[2] for managing and publishing e-journals. The system has been designed to be installed and run locally by journal editors with minimal technical skills and technology access. The system will support open-access, peer-reviewed publishing with international collaboration among editors, and offers precise metadata indexing of published materials on a global scale through the use of Open Archives Initiative standards.[3] This free system is intended to reduce the cost and raise the quality of publishing in ways that make open-access publishing a realistic alternative to the traditional model. Open Journal Systems also includes tools that support greater public and professional use of published research.[4]

OPEN JOURNAL SYSTEMS AND RESEARCH CAPACITY DEVELOPMENT: THE CASE OF KENYA

A research project is being carried out in Kenya focusing on possibilities of using new publishing technologies to enhance research capacity development through publishing and sharing of information related to research findings in the public sector, particularly in African universities and government departments. This research will

[1]See http://pkp.ubc.ca.
[2]See http://pkp.ubc.ca/ojs/.
[3]See http://pkp.ubc.ca/harvester.
[4]See http://pkp.ubc.ca/demos/rsttour/.

establish the prospects of greater participation in the international network for the creation and distribution of knowledge by exploring the prospects, processes, issues, and hopes involved in initiating an online scholarly journal. Specifically, the ongoing research project is investigating whether the Open Journal Systems can be used to boost research capacity through sharing of knowledge and information, locally and internationally. The ongoing study is guided by the following objectives:

- Establish from the perspective of the scholarly community—teaching staff, students, librarians, and policy makers—the potential for electronic journals to contribute to both information exchange and research capacity development.
- Explore the technical and economic feasibility of supporting electronic journals that can contribute to knowledge circulation and sharing among scholars and other stakeholders.
- Find out whether such an electronic journal can facilitate local knowledge creation within a global exchange of knowledge, as well as foster networking among peers.

The hypothesis is that from technical, economic, social, and intellectual perspectives, new publishing technologies can provide a means of improving Kenya's research capacity by contributing to local knowledge development, as well as to a larger global exchange of knowledge.

The expected outcomes of the research project are:

- an informed analysis of the challenges, possibilities, and obstacles in pursuing online scholarly publishing in African universities as a means of improving the universities' research capacities and of providing greater participation in global knowledge systems. This analysis will serve as a guide for establishing new online journals in different fields, as well as for building better publishing software.
- an informed body of scholarly researchers actively participating in knowledge creation, sharing, and dissemination, as well as professional debates and discussions in the relevant fields of knowledge.
- improved networking among faculty, professionals, and policy makers, emphasizing use of scholarly and local works in education.
- greater global visibility for African scholarly publishing, particularly Kenyan knowledge products and scholarly contributions. It is hoped that the e-journal will give the stakeholders the means and voice to be heard.

Online Publishing in Kenya

Although the research is in its initial stage, the findings are quite encouraging. A baseline survey has been done to find out whether there have been attempts to apply the new technologies for publishing and sharing research findings, identify the challenges and successes in this area, and establish possibilities of networking. There are various organizations in Kenya involved in online journal publishing. Currently several journals (about five) are publishing online with the help of Bioline International.[5] Their field of specialization is mainly the sciences—biotechnology, medicine, insect science, food technology, and nutrition. The host organizations receive manuscripts from authors in hard and electronic formats, edit them, and organize them for peer review. After corrections are made the editors of the various journals send them to Bioline for posting on the Bioline Web site, where the abstracts can be accessed freely. Access to full articles requires a subscription.

The *African Journal on Food, Agriculture, Nutrition and Development*, formerly known as the *African Journal of Food and Nutritional Sciences*, has struggled to publish in both print and online even with the support of well-wishers and friends. It was established in Kenya to provide a platform from which issues and scientific information concerning Africa could be effectively addressed and shared. The journal was established largely as a capacity-building initiative. The impetus was for professionals in nutrition, agriculture, and development, who

[5]For additional information on Bioline International, see Chapter 14 of these *Proceedings*, "Bioline International and the *Journal of Postgraduate Medicine*: A Collaborative Model of Open Access Publishing," by D. K. Sahu and Leslie Chan.

were native to or living in Africa, as well as non-Africans with an interest in Africa, to have an opportunity to come up with practical and sustainable solutions to the continent's problems. It would also encourage adoption of the writing culture among budding professionals in these fields. The title of the journal was changed because it was shared with two other journals; they also wanted a title that would reflect their main goals of linking research to development. The journal is peer-reviewed and has an editorial board, along with a secretariat composed of recent graduates from the university in the areas of food science and technology. It has its own Web site,[6] and publishes full articles in English and abstracts in both English and French. Access to full articles and abstracts is free. For translations into French the journal relies on friends.

The Kenyan research study has identified several benefits of online publishing for the journals surveyed. Online publication increases visibility for the journals, their authors, and the research findings and discussions. Researchers feel that the online journals have given them a voice to be heard outside their borders. Publishing online also increases readership and exposure to more sources of information through the Web site links. It has encouraged partnerships and collaboration among authors in related fields locally and internationally. Works can be published and accessed faster than before.

Several challenges facing online publications were also identified. Funding, especially for printing costs, is a big problem. Some journal editors interviewed saw no hope of surviving if alternative funding sources were not found. Distribution is another problem, for mailing costs are quite high. Sluggishness on the part of some editors, their assistants, and reviewers can lead to delays. The editorial and reviewing jobs were voluntary in all journals visited, and there appears to be a lack of incentive. There is also a lack of training for some reviewers, especially on scientific writing and research methodology. Other challenges identified include the poor means of communication and lack of access to technology. Readers in rural and small urban areas in Kenya cannot enjoy the benefits of online publishing systems directly. If print copies are not also available there is less readership of the journals. There was also a lack of awareness of the new publishing technologies and the way they work. Some journals may not understand the potential of these technologies, which in some cases are viewed as a threat to their survival.

The Kenyan journals found many ways to cope with the challenges. In terms of funding, most journals rely on the proceeds from subscriptions and advertisements to cover their operating costs. A few rely on the goodwill of friends, well-wishers, and donors. In terms of access to technology most of the journals depend on the sponsoring organizations to publish online. These organizations are located in the capital city, Nairobi, where communication is not a big issue. All of the organizations visited produce both print and online copies to ensure that the technologically disadvantaged get access to the journal. Journals rely on the efforts and networks of their editors-in-chief and the goodwill of their editorial boards to overcome editorial and review challenges. Some journals are organizing in-house training sessions to coach inexperienced editors and editorial assistants. Lastly, seminars and workshops on open access journal publishing are being conducted with assistance from donors to expose stakeholders to the new technologies.

The Situation at the Universities: The Case of Kenyatta University

A visit to Kenyatta University was made to observe and interview academic staff and librarians about developments in new technologies and library resources. During the visit it was found that the university had made great strides in embracing new technologies, and use of information and communication technologies was also increasing. The university has constituted an ICT board under the deputy vice-chancellor to oversee implementation of ICT policies. The library has a seat on the board. The university has been the recipient of many ICT-related donor projects that have boosted its ICT base (e.g., the World Bank's African Virtual University, the Institute of Continuing Education through distance learning, and the electronic Supply of Academic Publications (eSAP) project). The university is also the depository of World Bank publications, both print and online. Kenyatta University also has many Internet cafes where staff and students can access all the information they need, and the library is being automated and computers have been purchased to facilitate the process. The library is working on

---

[6]See http://www.ajfand.net.

a retrospective conversion of its catalogue. It also has a few computers connected to the Internet to which postgraduate students and staff have unlimited access.

Kenyatta University publishes the *East African Journal of Science: An International Journal of Pure and Applied Science, Chemchemi: International Journal of the School of Humanities*, and the *Kiswahili Journal.* The latter is published sporadically because journal publishing is seen as very demanding and less rewarding. As such, academic staff would rather engage in writing school textbooks that earn better money. Journal subscription has decreased since the 1980s due to reduced budgets among university libraries. The acquisition librarian said at times the budget could not purchase even five journals and felt that equipping the library to be the center of research capacity development was not given priority. However, it was noted that there is slight improvement in the 2003 financial year; budget allocation for journals has gone up to K Sh. 1.1 million (about US$14,000). Even with this improvement the library can only subscribe to core journals, at times only one per department. The limited budget requires the university to buy journals in single copies through an agent. Other journals are acquired through donations, though these have decreased over the years.

Online Resources. The university library's focus is to expand its electronic resources. The library has Internet connectivity and allows postgraduate students and academic staff unlimited access. It has access to electronic and online journals (close to 8,000) and bibliographic databases with abstracts, including Dissertation, Psychlit (psychology), TEEAL (environment and agriculture), POPLINE (population), EBSCO, Humanities Index, Education Index, Medline, and Elite. Most of these databases are available on CD-ROM. The library offers online publications by the International Network for the Availability of Scientific Publications' (INASP) Programme for the Enhancement of Research Information (PERI), including EBSCO online, Blackwell Synergy, AJOL, and the IDEAL library. It also offers electronic books, primarily World Bank publications. INASP/PERI has been helping in paying for the licenses of the databases.

Another new project in Kenyatta University is eSAP, whose mandate is to train library staff, faculty members from various departments on Internet, publishing, and Web design. The project also aims at facilitating electronic publishing of resources from the university, starting with business and development studies. The project is still at the infancy stage and has no editorial board. Initially the software was to be installed at Kenyatta University, but due to infrastructural (communication mainly) problems, inadequate articles to publish in the targeted disciplines and editing and reviewing problems, they shifted the base to the Netherlands.

Donors funding the Kenyatta University's electronic journal initiative include the World Bank, INASP/PERI, the Rockefeller Foundation, and Cornell University (host to the TEEAL databases).

Building research capacity through online publishing has encountered many challenges, including inadequate exposure to and training in computer applications, especially the Internet, even among the library and academic staff; inadequate infrastructure, including equipment (computers and their accessories are few), furniture, and communications (reliable telephone connections); lack of a technical support staff; and low morale of likely authors due to low salaries paid in an ever-rising cost-of-living environment. The dilemma for these potential authors is: devote time to research and writing or to looking for means of survival? The challenges related to editing and reviewing stem from the incentive structure related to these activities and the lack of financial resources. There is also a concern about the sustainability of the resources after the donor has left.

RECOMMENDATIONS REGARDING ONLINE PUBLISHING

There is a need to create an awareness of the Online Journal Systems project and its long-term benefits. There is also a need for capacity building for would-be editors and reviewers, especially in scientific writing and research methodology, even at the university level, as well as for technical staff and library staff. Budgets must accommodate the purchase and maintenance of equipment, especially computers and communication gadgets. There is an urgent need to look for ways to motivate authors, editors, and reviewers, such as training, experience-sharing workshops, and networking opportunities.

33

Metadata Clearinghouse and Open Access to Geographic Data in Namibia

Ndaendelao (Emma) Noongo and Nico Willemse
Ministry of Environment and Tourism, Namibia

This presentation focuses on the metadata clearinghouse and open access to geographical data in Namibia. Namibia is located in southwest Africa and comprises 825,000 square kilometers, approximately the size of France and Germany together. It has a population of 1.8 million, according to the 2001 census. The population is growing at a rate of 2.8 percent.

Namibia is an arid country; only 3 percent of the land is arable, 1 percent of it is in permanent crops, 46 percent is permanent pasture, and 22 percent is forest and woodland. About 17 percent is identified for conservation. The rest is wasted land. The climate of Namibia is typical of a desert country—very hot and dry, with variable distributed rainfall.

Namibia, after its independence in 1990, became one of the first countries in the world to include an environmental clause in its constitution. That is a luxury owing to the climate of the country, and given that the majority of the Namibian people rely on natural resources for their livelihood. It is very important to manage natural resources while caring for the environment. Therefore, Namibia must explore means of sustainable development.

DEVELOPING AN ENVIRONMENTAL INFORMATION SYSTEM IN NAMIBIA

In January 1998 the Ministry of Environment and Tourism, jointly with the Government of Finland, launched a 4-year (1998-2001) national program—Information and Communication for Sustainable Development (Infocom) with the aim to promote sustainable development in Namibia through

- developing an effective Environmental Information System unit within the Ministry of Environment and Tourism;
- developing communication mechanisms to disseminate environmental information; and
- defining sets of National Core Set of Environmental Indicators, and compiling the State of Environmental Indicators based on the former.

"We can only manage things we can measure"
To measure the environmental indicators, there is a great need for data.

Infocom started with defining the environmental indicators based on the thematic reporting of the state of the environment. For each indicator defined a data set was collected (or efforts were made) for monitoring purposes. This approach, nevertheless, had many shortcomings largely owing to a lack of communication in the scientific community. As the indicators were defined thematically, there was little communication among data-producing agencies as well as data users; the process of data collection allowed duplication of efforts. Thematic groups ended up collecting similar datasets, as some or many of the indicators were cross cutting the thematic domains.

It was not until Infocom's review and a stakeholder analysis in December 2000 that Infocom realized that it had to broaden its scope. Both the project's term review and the stakeholder analysis indicated that a lack of easily available, up-to-date, and reliable data was a big problem in environmental decision making in Namibia. The team therefore, altered its main project components and work plan, which was adopted by its Steering Committee in 2001. The following specific adjustments were made to the approach and workflow of the project, as the team strived to

1. make available and manage environmental data in the form of a metadatabase; and
2. initiate and facilitate communication with and among data-producing agencies and data users.

These two main adjustments marked the progress toward an operational spatial data infrastructure in Namibia. To date, an Environmental Information System unit is fully established and is incorporated in the ministry's structure.

The Environmental Information System unit seeks to provide infrastructure that supports the collection and maintenance of geographical information in Namibia. A significant aspect of this effort is to ensure complete and efficient access to data. Activities are currently being funded jointly by the governments of Namibia and Finland. This donor assistance will terminate at the end of 2003. To sustain these activities the Namibian government has incorporated the unit's activities in the structure of the Ministry of Environment and Tourism. When the project is phased out, the government will continue the activities.

EFFORTS AND ACTIVITIES TO MAKE DATA ACCESSIBLE IN NAMIBIA

Before discussing the accessibility and availability of geographic data it is imperative to consider Namibia's data access policy, which has been the subject of many discussions since independence in 1990. A number of resolutions were passed by professional bodies at the end of their seminars, workshops, and meetings; however this approach was neither sufficient nor appropriate for such important issues. There was no responsible body or institution mandated with the responsibility to develop a national data policy. The Environmental Information System unit was created to be an information center that would coordinate the efforts of various stakeholders in this area and to monitor data sharing and facilitation in Namibia.

The Environmental Information System unit established the Environmental Monitoring and Indicator Network in 2001. This network is made up of institutions and individuals that are involved in data collection or are using data in the ministries, NGOs, and in the private sector. Network officials strongly recommended that Namibia form a committee to be responsible for the national spatial data infrastructure development in the country. This committee is responsible for data sharing and exchange policy, data storage and facilitation, and standards and guidelines, including metadata standards. A cabinet paper for formulating the data sharing and exchange policy is currently under review. This paper addresses such issues as cost of the data, data formats, and intellectual property rights.

The Environmental Monitoring and Indicator Network is also looking at data standards that will allow the data to be compatible and interoperable for different uses. This is especially important because different institutions have developed data using their own standards. The data are not compatible and cannot be exchanged. Therefore, one has to duplicate effort and collect data that have already been collected, because the available data do not comply with their format. The Environmental Information System unit is developing standards regarding the format, content, classification, and exchange of data for implementing the national spatial data infrastructure. Eventually these standards will also align with the work of the international standards technical committee, once it is available.

The accessibility of data in Namibia is increasing, even though the draft data access policy is still under review.

In the past it was a problem for institutions to exchange data even if the data were in compatible formats. One could only get the data by knowing an official in the next institution; otherwise, it was very difficult. This has changed recently. Most institutions now show a willingness to distribute their data in digital format on request.

Over the past years the Environmental Information System unit has initiated efforts to produce and make digital databases freely available. A national *Atlas of Namibia* (hardcopy format) and a digital Atlas Database were completed toward the end of 2002. The atlas provides basic reference material on the geography of Namibia, including social, demographic, economic, infrastructural, physical, climatic, and biological features of the country. Throughout the book relationships between these features are explored to highlight the most important and interesting environmental potentials and constraints, especially as these relate to sustainable development options as Namibia enters the twenty-first century. The book has been designed and written in a format to reach as wide an audience as possible.

All data obtained for the production of the book are being made freely available wherever possible. This is being done as a service and in the spirit that all data should be as readily available as possible. All datasets are accompanied by metadata. Most of the data are in ArcView format and accessible as zipped files. The data are downloadable from the ministry's information portal.[1] Upon request the data are distributed on CDs to those who do not have access to the Internet.

A series of Regional Environmental Profiles have been completed, and more are being produced or planned. The profiles provide environmental information, compiled on a regional basis, for planning, managing, and monitoring of natural resources toward sustainable development.

The information compiled in the profiles is environmental in an encompassing sense. The profiles provide baseline information on such natural resources as water, soils, fauna, and flora. Essential information on climate, demography, agriculture, infrastructure, environmental threats and issues, land tenure, and governance is also provided. The information is gathered from various organizations collecting and using data from the government, NGOs, and the private sector, and from studies initiated by Environmental Information System projects both in the field and remotely, such as satellite images and aerial photographs. Wherever possible data are collected with geographical coordinates to allow the information to be mapped and used in spatial analyses.

To date, two regional profiles have been completed: the Caprivi Regional Profile (1997)[2] and a North Central Profile (2000).[3] The North Central Profile covers the four regions in the central north: Oshana, Omusati, Oshikoto, and Ohangwena. Natural resource mapping in the Kavango region has also been initiated for the third environmental profile, while the planning for a profile of Kunene region is envisaged for the near future. The profiles produce enormous digital databases and the data are available at no cost, under the principle that human development is stimulated by the unrestricted flow of information. The data can be downloaded from the ministry's environmental portal. As with the Atlas data, the data are distributed on CDs upon request to those who do not have access to the Internet. All data distributed via Internet or on CD ROM are accompanied by metadata.

Most of the data are available in unprojected latitudes and longitudes in ArcView shapefiles. Other data are available in Microsoft Excel files. Besides the digital data there are satellite images, aerial photographs, and bibliographies available through the Environmental Information System unit on the regions covered thus far.

Metadata have rapidly become a buzzword among data users in Namibia. The creation of metadata can be regarded as the essential point to describe data and improve on data availability and access. As a component of data infrastructure, metadata are key to maximizing community access to information and helping users find the information they need. The benefits of having quality and accessible metadata are clearly evident from the National Metadata Directory.[4] A metadata Web page has been created that includes introductory information on

[1] See http://www.dea.met.gov.na/data/Atlas/Atlas_web.htm.

[2] See http://www.dea.met.gov.na/data/Caprivi/caprivi_index.htm.

[3] See http://www.dea.met.gov.na/nnep/index.htm.

[4] See http://www.dea.met.gov.na/data/publications/Databases/MetaDB/metadataEnv.htm.

metadata, use, and management thereof. The Web page contains over 500 metadata records from various agencies involved in data collection and usage.

The Environmental Information System unit has developed a metadata standard that sets out minimum requirements for metadata to be included in the National Metadata Directory. The standardized metadata design is for use by data custodians to create, store, and distribute core metadata elements. The standard, which conforms to Federal Geographic Data Committee standards, has been widely adopted and used in the creation of metadata by various agencies. The Environmental Information System unit plans to ensure that the metadata standards comply with standards set for both the South African Development Community and the International Organization for Standardization's Technical Committee.

The Environmental Information System unit has also created and registered a metadata clearinghouse node for Namibia at the global Metadata Clearinghouse, where metadata can be searched and viewed at the international level.[5] The global clearinghouse activity, which is sponsored by the U.S. Federal Geographic Data Committee, is a decentralized system of servers accessible through the Internet that contain field-level descriptions of available digital spatial data. This clearinghouse allows individual agencies, consortia, or geographically defined communities to band together and promote their available data.

Over the past 19 months the Environmental Information System unit has made efforts to systematize and consolidate the resources in the different directorates' libraries. In November 2002 the Environmental Information System unit officially launched the Ministerial Resource Center, well known as the MET Resource Center, with the mission to "provide environmental related information to the Namibian nation at large, especially the scientific community and students."

A database has been created combining the resources from the Directorate of Environmental Affairs, the Directorate of Forestry, and the Etosha Ecological Institute in Okaukeujo. This database is being maintained and updated regularly by the main center at Environmental Affairs. It is available for viewing and information searches on the Resource Center Web page on the ministry's Web site.[6]

Our Web site is rich in information. Research discussion papers, policy, and legislation are also provided as downloadable files. Descriptions and contact details for ongoing projects in the ministry are also available, as well as links to other natural resource sites and a variety of geographic and textual resource data.

Worthwhile mentioning is the good working relationships that government agencies have with local NGOs. Not only are the government and its institutions providing data, but NGOs and private companies are also making data available on the Internet.[7] In the Environmental Monitoring and Indicator Network the NGOs and private companies are also well engaged, and some are members.

The information is available on national and local levels. There are local people working on the ground who are quite involved from the very beginning of the projects. They assist with the data collection. Once the data are captured into geographic information system environments, the maps are returned to the local participants for verification of the information. So the data reach the localities, and are not only at the national level.

Currently we do not sell the data. All data are distributed free of charge. Only a few government institutions have cost recovery mechanisms for providing the data. It is unique to Namibia that their NGOs and the government work so closely together; the government commissions them to produce data and the data come back to the government free of charge. So, in a way, sharing of the data is natural.

The success of this initiative involves political support, as well as coordination and cooperation. Initially there was no single body or individual responsible for creating or establishing infrastructure in Namibia. Making it work required collaboration and the production of a Web site, which hosts all of the environmental data.

---

[5]See http://clearinghouse4.fgdc.gov/registry/clearinghouse_sites.html.
[6]See http://www.dea.met.gov.na/wwwisis/MET.01/form.htm.
[7]See, for example, http://www.drf.org.na.

34

Open-Access Initiatives in India

T. B. Rajashekar
National Centre for Science Information, India

There are two components to this presentation. The first focuses on the relevance of open-access publishing in developing countries; the potential for open-access publishing in India; and a few current open-access initiatives in India. The second component proposes a possible technical model to organize open-access publishing in India.

OPEN-ACCESS PUBLISHING IN DEVELOPING COUNTRIES

Open-access publishing is the provision of free online access to quality scholarly material that can be defined as "open domain," meaning publicly supported research information, and "open access," so that it is copyrighted to be freely available scholarly material. Several enablers have motivated the development of open-access publishing. These include global movement and initiatives for open access like the Open Archives Initiative, the Budapest Open Access Initiative, Scholarly Publishing and Resources Coalition, and Free Online Scholarship, as well as the initiatives undertaken by ICSU, UNESCO, and CODATA.

Other important enablers are availability of free online publishing and digital repository management software and protocols for metadata standards. Using these standard protocols for metadata allows interoperability among the repositories for sharing information and providing centralized services.

Much has been said about the value of open-access publishing in developing countries. Open-access publishing enables researchers in developing countries to establish priority for their research, which they could use later to defend their intellectual property. It removes excess barriers in terms of both price and permission, enhances national research capacity, and improves visibility for developing-country research. Open access thus enables a global platform for this research and collaboration and reciprocates the information flow from South to North among all countries. It is hoped this also leads to improved economy.

OPEN-ACCESS INITIATIVES IN INDIA

In India, there is a large opportunity for open-access publishing. There are many noncommercial research and development institutions, both academic and research laboratories. For example, there are approximately 300 universities that offer both graduate and research programs. There are also many R&D laboratories operating within government science agencies, which cover domains like industrial research, defense research, agricultural research, medicine, ecology, environment, information technology, space, energy, and ocean development. These

institutions, which produce research work, could potentially convert their data into online accessible material. Many of these institutions, and also several professional societies, publish science journals. Tools like the Open Journal Systems[1] could help many of these journals to come online in an open-access environment.

Technical reports produced by many R&D projects, laboratories, and other institutions would also be candidates for providing open access. Theses and dissertations at universities, conferences, research papers—whether preprints or postprints, unpublished research findings, data, or standards are candidates for open-access publishing in India and in other countries.

The following examples of open-access initiatives in India are drawn from scholarly science journals, theses, institutional archives, books, data and open access at the metadata level, and open access at portal and gateway services.

One example is the Indian Academy of Sciences, established in 1934. The Indian Academy of Sciences is one of three science academies in India. Apart from various other activities it publishes 11 science journals reporting research work both in India and outside. These journals, mainly in print, are freely accessible on the Web.[2] The Indian Academy of Sciences is currently digitizing all the archival issues and expects to post them online very soon. The managing editor of these journals noted that offering these journals on the Web has increased subscriptions to the print journals from foreign countries, because more researchers and libraries outside India are learning about them.

The Indian National Science Academy publishes journals, proceedings, and monographs and provides these online.[3] Vidyanidhi, meaning the "treasure of knowledge" in Sanskrit, is another open-access initiative that is trying to digitize and host theses and dissertations. It operates from the University of Mysore and is part of the global electronic thesis and dissertation initiative.[4] The Vidyanidhi project is also developing workflows and definitions and addresses multilingual support issues. Other institutions are putting their theses online as well, including the Indian Institute of Technology in Delhi.

The E-print archives of the Indian Institute of Science[5] is an online digital repository of research papers, both preprints and postprints, technical reports, unpublished findings, and journal articles of the faculty. It was set up using eprint.org open-source software, and is registered in the e-prints registry. Eprints@iisc is now part of the worldwide institutional e-print archives. The E-prints archives allow the faculty and students to submit their publications electronically to the campus network. Although depositing is not allowed from outside the campus, access is allowed from anywhere on the Internet. The eprints@iisc Web site also supports metadata for browsing and searching. It is also integrated with the Greenstone Digital Library software, which enables full-text searching of the e-prints.

The Universal Library is another interesting project. It is funded by the Office of the Principal Scientific Advisor to the Government of India and is hosted by the Indian Institute of Science in collaboration with the Carnegie Mellon University in the United States.[6] The goal of this project is to provide a free, searchable collection of 1 million books that are no longer copyrighted. The collection is also expected to act as a test bed for research in language processing, indexing, and retrieval.

There are some examples of initiatives that provide open access to data as well. The National Chemical Laboratory is a national research lab in India that provides free access to their data, including data from the National Collection of Industrial Micro-organisms and the National Centre for Biodiversity Informatics.[7] There are also open-access initiatives at the metadata level. INDMED, at the National Informatics Centre in Delhi, is a bibliographic database of Indian biomedical literature and indexes 75 Indian journals.[8] There is also a backup

---

[1]For more information on the Open Journal Systems, see Chapter 32 of these *Proceedings*, "The Public Knowledge Project's Open Journal Systems," by Florence Muinde.

[2]See http://www.ias.ac.in. This site offers links to the various science journals published by the Indian Academy of Sciences.

[3]See http://insa.ac.in.

[4]See http://www.vidyanidhi.org.in.

[5]See http://eprints.iisc.ernet.in.

[6]See http://www.dli.ernet.in/collections/Books/Books.html?path1=collections%2FBooks.

[7]See http://www.ncl-india.org/ncim and http://www.ncbi.org.in, respectively.

[8]For additional information on INDMED, see http://indmed.nic.in.

document delivery service associated with this. The University Grants Commission is the body that coordinates all Indian university education. It supports an information library network program INFLIBNET that makes meta databases related to R&D projects available on the Web.[9]

There are other major open-access initiatives as well. For example, the World Health Organization and the Indian Council for Medical Research are working together on the National Health Information Collaboration. The project provides a portal for Indian health data and information and free open software, which enables people to use it for other purposes. The Council of Scientific and Industrial Research is responsible for scientific industrial research in India and has a unit for R&D for information that aims to provide open access to Indian patents and medicinal plants information.

There are also interesting gateway services that integrate access to other open-access resources on the Internet. SciGate at the Indian Institute of Science is a science information portal that integrates a variety of science information sources on the Web. Another example is AeroInfo at the National Aerospace Laboratory, which provides an aerospace virtual library.

AN OPEN-ACCESS PUBLISHING MODEL FOR INDIA

How can these examples of open access be replicated and adapted in an organized manner across India? One proposal is a national network of distributed, interoperable, open-access digital repositories of research material, both at the institutional level and across the institutions in open-access science journals and conferences. The motivation for this network is the strong support for open-source software in India and the increasing interest to use digital library software, such as the Greenstone library software developed by the New Zealand Digital Library Group. This software has been used innovatively to publish content, both on the Web and on CD. There is an emerging model provided by the E-print archives, using the Open Archives Initiative interoperability framework, which makes this software compliant with that initiative.

How do we realize this model? Academic institutions can set up institutional repositories of their research output. Science journals can adopt open-access publishing that is compliant with the Open Archives Initiative. New online open-access journals that focus on areas of local strength, such as agriculture and medicine, and graduate student journals could also be established.

A key issue is the incorporation of peer review and quality control in such an open-access publishing environment. Peer review must be employed at the institutional level or across the institutions to ensure that we put quality materials on these repositories. As of now, however, there appears to be no consensus on an effective mechanism for establishing quality control in such systems.

How do we operate this model? These repositories at the institutional levels act as data providers and provide metadata for harvesting. This is possible because of the availability of software that is compliant with the Open Archives Initiative (e.g., eprints.org, DSpace, CERN CDSWare, and Open Journal Systems). Some of these institutions themselves could act as the service providers by harvesting the metadata from different repositories and offering a variety of national level services. One such service could be a central metadata index service, wherein a user could identify papers, and then go to the actual research paper hosted on an institutional repository.

Libraries have a major role to play in this. For example, they could lead the way in establishing and operating an institutional repository and supporting researchers with open-access publishing activities. Libraries are best suited to provide document preparation and content management expertise.

CONCLUSION

A national-level mechanism is essential to promote and coordinate open-access publishing systems and to improve awareness for open access. Training is also very important, in terms of tools, processes, and standards.

---

[9]See http://web.inflibnet.ac.in.

There should be wide support for setting up working models and services. National resource centers for open-access publishing in developing countries could lead the way in setting up a working model. Once the system starts evolving, it could run on its own.

It is also very important to have the support of organizations like UNESCO, ICSTI, ICSU, and CODATA to promote and support these initiatives.

35

Closing Remarks

M. G. K. Menon
LEAD, India

The purpose of this symposium was to discuss the area of public domain and the question of open access to data. There has been a great deal of clarification on a wide range of issues through the discussions that have taken place. For example, many presentations addressed the economic aspects of open access. Open-source software is very important and will be of great value. However, one should not minimize the task involved. It is going to be expensive, complex, and it will take time. It will have particular relevance to developing countries that cannot afford to pay repeatedly for proprietary software, as systems are upgraded and earlier ones are rendered obsolete. The focus on open access, public domain, and information needed for science laid particular emphasis on the needs of developing countries.

Presentations on the legal aspects of open access examined current policies in this area in the United States and in Europe, and also with reference to developing countries. What came out was a scenario that was highly varied, with some countries being more restrictive and others more open.

Apart from a general discussion on these issues a large number of interesting examples and field demonstrations were highlighted, illustrating how initiatives involving the North and the South can overcome problems that lead to the digital divide.

Governments that are the participants in intergovernmental meetings and that therefore help to produce the World Trade Organization, the Trade-Related Aspects of Intellectual Property Rights Agreement, the World Intellectual Property Organization, and European database directive policies do so on the basis of self-interest. Commercial interests lobby for what they require. Science however is much more diffuse. Science does not lobby per se with governments. The needs of science are, however, something that governments must be made to understand, whether this be in terms of access to data or on a more general basis an ambience of openness, freedom, excitement, transparency, and integrity.

The entire basis for the developments that have enabled today's information society has come from work in science, whether it was the Internet, the World Wide Web, the developments in computer science, including hardware and software, the mathematical underpinning, solid-state electronics (and more particularly micro-electronics), lasers, wireless, space technologies, and much more. This impact of science is somewhat like the large part of an iceberg below the surface of the sea—it has to be conveyed to policy makers and understood by them. Because science does not lobby for its interests, its case often gets left behind or forgotten.

To some extent when we debate such issues at forums such as this symposium, the participants represent a rather coherent, homogeneous community with interests in and an understanding of science. The problems we

encounter are with external driving forces, and they are not usually present. It is very important that we do not just preach to the converted but also look at what we need to do to make the needs of science appreciated by governments, business, industry, and more generally, by society.

The World Summit on the Information Society (WSIS) presents an opportunity for such outreach. CERN plans to convene a preparatory meeting before the first WSIS to strategize on making the case for science. We need to participate in this and give it our fullest support. We have a good case, but we have to make it.

APPENDIXES

Appendix A

International Symposium on Open Access and the Public Domain in Digital Data and Information for Science

UNESCO Headquarters
Fontenoy Room II
Paris, France
March 10-11, 2003

Jointly organized by

International Council for Science (ICSU)
United Nations Educational, Scientific and Cultural Organization (UNESCO)
The U.S. National Academies
Committee on Data for Science and Technology (CODATA)
and
International Council for Scientific and Technical Information (ICSTI)

PROGRAM

Monday, 10 March

8:00 Registration

8:40 Welcome by CODATA President
Shuichi IWATA, University of Tokyo, Japan

8:45 Introduction by Symposium Chair
M. G. K. MENON, LEAD India

8:55 Keynote Address
David DICKSON, Science and Development Network, United Kingdom

Session One—Legal, Economic, and Technological Framework for Open Access and the Public Domain in Digital Data and Information for Science

9:20 Introductory Remarks by Session Chair
Elizabeth LONGWORTH, Industry New Zealand

9:30 Overview of Legal Aspects in the European Union
Thomas DREIER, University of Karlsruhe, Germany

9:55 The Legal Context in Developing Countries
Alan STORY, University of Kent Law School, United Kingdom

10:20 Economic Overview of Open Access and the Public Domain in Digital Scientific and Technical Information
Robin COWAN, MERIT/University of Maastricht, The Netherlands

10:45 COFFEE

11:10 UNESCO's Approach to Open-Access and Public-Domain Information
Koïchiro MATSUURA, Director-General, UNESCO

11:25 Economic Considerations for Open-Access and Public-Domain Availability of Scientific Information in Developing Countries
Clemente FORERO-PINEDA, University of Bogota, Colombia

11:50 Information Technology and Data in the Context of Developing Countries
Chrisanthi AVGEROU, London School of Economics, United Kingdom

12:15 Discussion

12:45 LUNCH

The Opportunities and Challenges of Open-Access and Public-Domain Scientific Information in Developing Countries

Session Two—Data and Information in the Public Health Sector

14:00 Introductory Remarks by Session Chair
Dialo DIOP, Université Cheihk Anta Diop, Senegal

14:10 The Ptolemy Project: Delivering Electronic Health Information in East Africa
Massey BEVERIDGE, University of Toronto, Canada

14:35 Health Information for Disaster Preparedness in Latin America
Jean Luc PONCELET, Pan American Health Organization, United States

15:00 Bioline International and the *Journal of Postgraduate Medicine:* A Collaborative Model of Open-Access Publishing
D.K. SAHU, JPM Managing Editor, India, and Leslie CHAN, Bioline, Canada

15:25 COFFEE

Session Three—Data and Information in the Environmental Sector

15:50 Introductory Remarks by Session Chair
Farouk EL BAZ, Boston University, United States

16:00 Geospatial Information in Development
Mukund RAO, Indian Space Research Organisation

16:25 A Comparative Analysis of Data Access Policies in Meteorology
Peter WEISS, U.S. National Weather Service

16:50 Recent Developments in Environmental Data Access Policies in the People's Republic of China
LIU Chuang, Chinese Academy of Sciences

17:15 Discussion

17:50 ADJOURN

18:00 RECEPTION

Tuesday, 11 March

8:15 Registration

Session Four—Basic Sciences and Higher Education

8:50 Introductory Remarks by Session Chair
Lulama MAKHUBELA, National Research Foundation, South Africa

9:00 Information Needs for Basic Research: An African Perspective
Andrew KANIKI, National Research Foundation, South Africa

9:25 International Transfer of Information in the Physical Sciences
R. Stephen BERRY, University of Chicago, United States

9:50 Access to Scientific Information: Distance Education—The Ukraine and Other CIS Countries' Perspectives
Mikhail ZGUROVSKY, National Technical University of Ukraine

10:15 COFFEE

Session Five—Innovative Models for Public-Domain Production of and Open Access to Scientific and Technical Data and Information

10:40 Introductory Remarks by Session Chair
Dominique FORAY, Centre National de la Recherche Scientifique, France

10:50 A Contractually Reconstructed Research Commons for Scientific Data: International Considerations
Jerome REICHMAN, Duke University Law School, United States;
Paul UHLIR, The National Academies, United States

11:10 The Open-Source Paradigm and the Production of Scientific Information: A Future Vision and Implications for Developing Countries
Charles SCHWEIK, University of Massachusetts, United States

11:35 New and Changing Scientific Publication Practices Due to Open-Access Publication Initiatives
Erik SANDEWALL, Linköping University, Sweden

12:00 Overview of Open-Access and Public-Commons Initiatives in the United States
Harlan ONSRUD, University of Maine, United States

12:25 Discussion

12:55 LUNCH

Session Six—Examples of New Initiatives in Developing Countries

14:00 Introductory Comments by Session Chair
Alexei GVISHIANI, United Institute of Physics of the Earth, Russian Academy of Sciences

14:15 Overview of Initiatives in the Developing World
Sarah DURRANT, International Network for the Availability of Scientific Publications, United Kingdom

14:40 Open-Source GIS Software in Brazil
Gilberto CAMARA, National Institute for Space Research, Brazil

15:00 Public Knowledge Project Open Journal Systems
Sal MUTHAYAN, Doctoral Candidate, South Africa;
Florence MUINDE, UNESCO Fellow, Kenya

15:20 Metadata Clearinghouse and Open Access to Geographic Data in Namibia
Emma NOONGO and Nico WILLEMSE, Ministry of Environment and Tourism, Namibia

15:40 Open-Access Initiatives in India
T. B. RAJASHEKAR, National Centre for Science Information, India

16:00 General discussion

16:20 Closing Remarks by Symposium Chair
M. G. K. MENON, LEAD, India

16:30 ADJOURN

Appendix B

Biographical Summaries of Symposium Speakers and Steering Committee Members

SYMPOSIUM SPEAKERS

Chrisanthi Avgerou is a senior lecturer in information systems at the London School of Economics and Political Science. She teaches postgraduate courses on information systems implementation and on information systems in developing countries. She chaired the IFIP WG 9.4 on "Social Implications of Computers in Developing Countries" from 1996 to 2002. Her research is concerned with the study of the dual process of the utilization of information technology and organizational change within different socio-organizational contexts; one of her recent publications is on this subject (*Information Systems and Global Diversity,* 2002, Oxford University Press).

R. Stephen Berry has been the James Franck Distinguished Service Professor Emeritus in the Department of Chemistry at the James Franck Institute at the University of Chicago since 1989. He has A.B., A.M., and Ph.D. degrees and has taught at Harvard, the University of Michigan, and Yale. He has received a MacArthur Prize Fellowship and has spent extended periods at the University of Copenhagen, Oxford, Université de Paris-Sud, and the Frei Universität Berlin. Dr. Berry's research interests include the electronic structure of atoms and molecules; photo- and collisional detachment of negative ions; photochemistry of reactive organic molecules; vibronic coupling processes, such as autoionization, predissociation, and internal vibrational relaxation; thermodynamics of finite-time processes; dynamics and structure of atomic and molecular clusters; phase changes in very small systems; chaos and ergodicity in few-body systems; and most recently as an outgrowth of the cluster studies, dynamics on many-dimensional potential surfaces and the origins of protein folding. He has also worked extensively with the efficient use of environmental energy and other resources. Dr. Berry is also interested in issues of science and the law, and management of scientific data, activities that have brought him into the arena of electronic media for scientific information and issues of intellectual property in that context. He is a member and the home secretary of the National Academy of Sciences. He has been involved in many activities of the National Academies, including chairing the committee that produced the book *Bits of Power: Issues in Global Access to Scientific Data.*

Massey Beveridge is the director of the Office of International Surgery at the University of Toronto and is an attending surgeon at the Ross Tilley Burn Centre at Sunnybrook and Women's College Health Sciences Centre in Toronto. He has a longstanding interest in development issues, and was trained as an anthropologist at McGill and Cambridge universities. He has worked as a surgeon and surgical educator in a number of countries and now

focuses on surgical education and research for development. Dr. Beveridge earned B.A., M. Phil., and M.D. degrees and a Diploma in Tropical Medicine and Hygiene, and is a fellow of the Royal College of Physicians and Surgeons of Canada.

Gilberto Câmara is the director for earth observation at Brazil's National Institute for Space Research (INPE). He holds a B.S.E.E. from the Aeronautics Technological Institute, São José dos Campos, and an Ms.C. and a Ph.D. in computer science from INPE. He has written some 100 scientific papers and four books on design of geographical information systems, spatial databases, spatial analysis, and remotely sensed image processing. He is a consultant to the most important Brazilian funding agencies and teaches and supervises graduate students in INPE's graduate programs in remote sensing and computer science.

Leslie Chan is an associate director of Bioline International, a not-for-profit electronic publishing collaborative designed to improve global access to research published in developing countries. A trustee of the Electronic Publishing Trust for Development since 1997, Dr. Chan is active in promoting partnerships between educational and research institutions in the hope of narrowing the knowledge gap between the South and the North. He is one of the original drafters and signators of the Budapest Open Access Initiative, a worldwide movement that encourages open institutional archiving and free access to scholarly publications. As a lecturer in the Division of Social Sciences, University of Toronto at Scarborough, Dr. Chan teaches courses in new media, civic engagement, and international communications. He is frequently invited by international organizations to conduct workshops on knowledge management, electronic publishing, and instructional technology.

Liu Chuang is the director of the Global Change Information and Research Center, within the Institute of Geography and Natural Resources of the Chinese Academy of Sciences. She serves as the cochair of the CODATA Task Group on Preservation and Archiving of Scientific and Technical Data in Developing Countries, member of the ICSU Task Group on World Data Centers, associate director of the Data Committee of China Association of Geographical Information Systems, secretary general of Remote Sensing and Data Information Network, and member of the Chinese National Committee of International Geosphere-Biosphere Program. Dr. Liu Chuang was an information scientist and China Project Leader at the Consortium for International Earth Science Information Network (CIESIN) from 1994 to 1998; visiting professor of University of British Columbia, Canada, from 1992 to 1993; and associate professor of Peking University from 1989 to 1991. She served for UNDP/FAO and Asia Development Bank as a consultant and technical assistant during 1995 to 1998. She received her Ph.D. in geography from Peking University, China, and her master and bachelor degrees in geography in China. She has received awards from START (USA), CIESIN (USA), MOST (China), Peking University (China), and ISPRS (Japan) based on her scientific achievements.

Robin Cowan is professor of the economics of technical change at Maastricht University. He began his official affiliation with MERIT in 1996 as a professorial fellow. He studied at Queen's University in Canada and at Stanford University where he received a Ph.D. in economics and an M.A. in philosophy. Robin Cowan was assistant professor of economics at the University of Western Ontario until 1998. His current research focuses on technology competitions and standardization, the dynamics of consumption, and the economics of networks. He is also doing research on the changing nature of the economics of knowledge and intellectual property rights in the new economy. In the past he has done consulting research for the OECD on the economics of standards and the National Renewable Energy Laboratory on technological lock-in and renewable energy technologies. Recently, he has completed two research projects on "Intellectual Property Rights in a Knowledge-Based Economy" with Elad Harison for the Dutch Advisory Council for Science and Technology Policy. Professor Cowan is also an adjunct professor at the Economics Department at the University of Waterloo, Ontario, Canada.

David Dickson is the founding director of the Science and Development Network (SciDev.Net), a Web-based news and information service set up in 2001 that covers science, technology, and the developing world. A graduate in mathematics from the University of Cambridge, he has formerly been science correspondent of *The Times Higher Education Settlement*, the Washington correspondent of *Nature*, the European correspondent of

Science, and the editor of *New Scientist*. Prior to entering journalism, he was the first executive secretary of the British Society for Social Responsibility in Science. He is the author of *Alternative Technology* (London 1973), and *The New Politics of Science* (University of Chicago Press, 1986).

Thomas Dreier is a professor of law at the University of Karlsruhe, Germany, where he is the director of the Institute for Information Law. In the spring of 2002, Professor Dreier was Global Visiting Professor of Law at the New York University School of Law. Before joining the University of Karlsruhe, Professor Dreier worked at the Max-Planck-Institute for Foreign and International Patent, Copyright and Competition Law in Munich, Germany. He is vice-president of the Association Littéraire et Artistique Internationale and vice-chairman of its German national group, as well as a member of the Legal Advisory Board of the EU's DG Information Society and of the Advisory Panel on Intellectual Property of the Steering Committee of the Mass Media of the Council of Europe. Professor Dreier also acts as executive secretary of the German Computer Law Society (Deutsche Gesellschaft für Recht und Informatik, DGRI). He earned a J.D. (Munich) and a M.C.J. (NYU).

Sarah Durrant is a senior programme manager for the International Network for the Availability of Scientific Publications (INASP) Programme for the Enhancement of Research Information (PERI). She works with publishers and information providers to develop differentially priced access terms in order to make their products and services affordable, and therefore sustainable, in developing countries. Before joining INASP she had a 12-year career in international scientific, technical, and medical publishing at Harcourt Brace/Academic Press (now Elsevier) in London and John Wiley & Sons, working mostly on journals, particularly electronic journals. As manager of STM Journals Development for Wiley Europe, Sarah helped developed Wiley's online journals service, *Wiley InterScience*; this was followed by two years with the dynamic and innovative journals digitization and hosting company CatchWord/Ingenta.

Clemente Forero-Pineda is a professor at the University of the Andes and Rosario in Bogotá, Colombia. He earned a certificate in mathematics and physics and an engineering degree at Institut National des Sciences Appliquées de Lyon (France). He completed an M.A. and a Ph.D. in economics at Stanford University. He was dean of the School of Economics at Colombia's National University and director-general of Colombia's National Science Fund (Colciencias).

Andrew Kaniki is the executive director of Knowledge Management and Strategy at the National Research Foundation in South Africa. Until recently he was a professor of information studies; during the last three years he has been the pro vice-chancellor and acting deputy vice-chancellor (academic) at the University of Natal, South Africa. He has also taught at the University of Zambia and worked as a science information specialist at the Engineering and Science Library, Carnegie Mellon University in the United States. He holds a bachelor's degree in politics and library science (University of Zambia), an M.S. from the University of Illinois, and a Ph.D. and agricultural information specialist certificate from the University of Pittsburgh. He has published and presented several conference papers on information needs and use.

Koïchiro Matsuura was elected the eighth director-general of UNESCO on November 15, 1999. His studies in economy and law, which he started at the University of Tokyo and continued in the United States, permitted him to start a diplomatic career at a very early age. He entered the Japanese Ministry of Foreign Affairs at a young age as the third secretary of the Embassy of Japan in Ghana and West Africa. After a brief stint holding various positions at the central administration at the Ministry of Foreign Affairs in Japan, Mr. Matsuura then gained significant intergovernmental organization experience as second and then first secretary of the Japanese Delegation to the OECD. Stops as counsellor of the Embassy of Japan in the United States and consul general of Japan in Hong Kong proved essential to being named director-general of the Economic Cooperation Bureau within the Japanese Ministry of Foreign Affairs. While director-general of the North American Affairs Bureau, Mr. Matsuura began his formal writing career and has had numerous titles published. An accomplished author in the fields of economic cooperation, bilateral relations, and perspectives on development, Mr. Matsuura then represented Japan

at the 1994 G-7 summit as deputy minister for Foreign Affairs. From 1994 to 1999, Mr. Matsuura was called upon to serve as Ambassador of Japan to France and concurrently to Andorra and Djibouti during which he published Japanese Diplomacy at the Dawn of the 21st Century. His first contact with UNESCO was as chairperson of the World Heritage Committee of UNESCO where he showed a natural aptitude for dealing with cultural heritage issues.

Florence Muinde is a senior programmes officer for the Kenya Civil Service Reform Secretariat and a UNESCO/ Keizo Obuchi research fellow at the University of British Columbia, Canada, in the area of information and communication technologies and research capacity development in developing countries. She has a master's degree in education (library and information sciences) from Kenyatta University. She has taught in various secondary schools in Kenya and has worked as a trainer, researcher, and officer in charge of data management in the textbooks for a primary schools project sponsored by the U.K. Department for International Development and the Government of Kenya. In 2000 she won a scholarship to participate in a course at the University of Hamburg in information as a social resource. This culminated in the initiation of a virtual women's university and writing of a book titled *The Feminist Challenges in the Information Age.*

Saloshini Muthayan holds a doctoral fellowship from the National Research Foundation, South Africa and is a doctoral candidate at the University of British Columbia; she is also a research assistant with the Public Knowledge Project at the University of British Columbia (http://pkp.ubc.ca). Ms. Muthayan's research focuses on ways of building the research capacity of universities in the developing world, using South African universities as a case in point. In particular, she is examining how new open-access publishing technologies, such as Open Journal Systems, may assist in stanching the gap in knowledge access, production, and dissemination between universities in the developed and developing world. Ms. Muthayan has worked in several research projects and has led a research team in the evaluation and rationalization of the colleges of education in the Eastern Cape, South Africa. She holds a master's degree from Rhodes University and a B.A. with majors in international politics and English from the University of the Witwatersrand, South Africa.

Ndaendelao (Emma) Noongo is a database manager for the Environmental Information Systems Unit of the Ministry of Environment and Tourism. She was employed for two years as a researcher at Namibia's National Programme to Combat Desertification. Her research interests include geographical information systems, remote sensing, and data management issues, particularly in the developing regions. She is also the chair of a committee that aims to build data-sharing policy, data quality guidelines, and spatial data infrastructure in Namibia. Ms. Noongo received a B.Sc. in natural science from the University of Namibia, an M.Sc. in geographical information systems and data management from Durham University (U.K.), and is a Ph.D. candidate in geographical information systems and remote sensing at Joensuu University (Finland).

Harlan J. Onsrud is a professor of spatial information science and engineering at the University of Maine. His research focuses on the analysis of legal, ethical, and institutional issues affecting the creation and use of digital spatial databases and the assessment of the social impacts of spatial technologies. He teaches courses in information systems law, cadastral and land information systems, environmental law, and information ethics. Professor Onsrud is the chair of the U.S. National Committee on Data for Science and Technology and currently serves on the Mapping Science Committee of the National Research Council. A licensed engineer and attorney, he is a cochair of the Global Spatial Data Infrastructure Legal and Economics Working Group and is immediate past-president of the University Consortium for Geographic Information Science. He recently stepped down as editor-in-chief of the journal of the Urban and Regional Information Systems Association (*URISA Journal*) and has published in numerous engineering, geographic information systems, and legal journals.

Jean Luc Poncelet is a Pan American Health Organization (PAHO)/World Health Organization area manager in Emergency Preparedness and Disaster Relief for Latin America and the Caribbean, based in Washington, D.C. Since 1986, he has served as a disaster preparedness subregional advisor in San Jose, Costa Rica; Bridgetown,

Barbados; St. John's, Antigua; and Quito, Ecuador. Since joining PAHO in 1986, Dr. Poncelet has participated in major disaster response and humanitarian operations in the Americas and was among the first professionals to actively develop the Supply Information Management System that is now endorsed by the principal UN agencies and by governments. He also helped to establish and strengthen the Latin American and Caribbean Disaster Information Center based in Costa Rica. He has participated in the elaboration of several technical documents related to disaster preparedness, response, and mitigation. He holds a M.D. from the Louvain University, Belgium, and a master's degree in Public Health degree from Universite Libre de Bruxelles, Belgium.

T.B. Rajashekar is a principal research scientist at the National Centre for Science Information (NCSI) at the Indian Institute of Science (IISc), Bangalore. He is also the Centre's associate chairman. He holds an associateship in information science from Indian Statistical Institute and a Ph.D. in library and information science from Pune University. At NCSI, his responsibilities include development and management of variety of network-based e-information services for the IISc research community. He has guided the development of SciGate—the IISc Science Information Portal, E-JIS—the IISc e-journal gateway and ePrints@iisc-the IISc eprints archive service. He also teaches in NCSI's post-graduate training program on Information and Knowledge Management. He has taught and coordinated several national-level workshops on the Internet, search engines, digital libraries, and content management. He was a UNDP fellow at the University of Massachusetts, Amherst, during 1992 to 1993 and a visiting scientist at Informatics India Ltd., during 2000. He delivered the prestigious Prof. S.R. Ranaganathan Memorial Lectures in 2001. He has held R&D project grants from several government and private agencies. Some of the projects he has handled include the establishment of LIS-FORM, a discussion forum for L&I services in India; development of K-Library for ICICI-Knowledge Park; relevance ranking of CD-ROM search results; and development of a resource base for public domain software in computer networks and databases and their copyright information. He is a fellow of the Society for Information Science. He is also member of several national and international professional associations. His teaching and research interests include information and knowledge organization, information retrieval, and digital libraries.

Mukund Rao is the deputy director (Technology & Systems) for the Earth Observations (EO) System Programme Office in the Indian Space Research Organisation (ISRO) Headquarters, Bangalore. He is also involved in the systems and technology assessment/coordination for the Indian EO programme. His research and work experience is in the field of spatial information systems. He has been the key person in the design and definition of the Natural Resources Management System (NRIS) programme of ISRO/DOS and has completed a number of NRIS projects in support of district planning, urban planning, wasteland development, etc. As programme manager of NRIS, he has been involved in furthering the scope and concept of NRIS as a natural resources information repository. Presently, he is the key design person for the National Spatial Data Infrastructure and is associated in its strategy development, action plan development and implementation. Mr. Rao is the chair of CEOS Working Group on Education and Training; president (elect) of the Global Spatial Data Infrastructure; vice president of the International Astronautical Federation (IAF) and vice chair of IAF's EO Committee; member of IAF's Education Committee, IAF's CLIODN Committee; and member-secretary of the Centre for Space Science and Technology Education in Asia And the Pacific Advisory Committee and participates in many other international fora. Mr. Rao is the recipient of the Hari-Om Ashrams Vikram Sarabhai Young Scientists Award. He has an M.S. in geology from Gujarat University, a master's in philosophy in remote sensing, and is presently pursuing his Ph.D. in systems design of urban information systems.

Jerome H. Reichman is the Bunyan A. Womble Professor of Law at Duke University, where he teaches in the field of contracts and intellectual property. Before coming to Duke he taught at Vanderbilt, Michigan, Florida, and Ohio State universities and at the University of Rome, Italy. He graduated from the University of Chicago (B.A.) and attended Yale Law School, where he received his J.D. degree. Professor Reichman has written and lectured widely on diverse aspects of intellectual property law, including comparative and international intellectual property law and the connections between intellectual property and international trade law. Other recent writings have focused on intellectual property rights in data, the appropriate contractual regime for online delivery of computer

programs and other information goods, and new ways to stimulate investment in subpatentable innovation without impoverishing the public domain. Professor Reichman serves as a consultant to the U.S. National Committee for CODATA at the National Academies on the subject of legal protection for databases. He also is an academic advisor to the American Committee for Interoperable Systems; a consultant to the Technology Program of the U.N. Conference on Trade and Development; and was a consultant to the U.N. Development Programme's flagship project concerning "Innovation, Culture, Biogenetic Resources, and Traditional Knowledge."

D. K. Sahu is a consultant pediatrician and neonatologist. After completing his graduation he founded his own consulting company to guide students in their journey through medical education. He received his doctorate degree in pediatrics from Mumbai University. Presently he is working as the managing editor of the *Journal of Postgraduate Medicine* and is on the editorial boards of several professional journals. He serves on the Board of Trustees of the Prof. B. A. Bharucha Foundation. He is also a member of numerous professional societies and a fellow of the College of Physician and Surgeons. He is a specialist on the use of information technology in biomedical publishing, online management, and communication. As an independent consultant he has helped many journals in their management.

Charlie Schweik is an assistant professor in the Department of Natural Resources Conservation and the Center for Public Policy and Administration at the University of Massachusetts, Amherst. He has three primary areas of interest: environmental policy and management, public information technology and digital government, and the intersection of environmental management and information technology. He has a Ph.D. in public policy from Indiana University, a master's in public administration from Syracuse University, and an undergraduate degree in computer science. Before academia he worked as a programmer at IBM and as a consultant to the U.S. Department of Energy.

Alan Story teaches intellectual property at Kent Law School in Canterbury in the United Kingdom. A Canadian, he was an investigative and political journalist with the *Toronto Star* (Canada) before making a career change. His earlier research looked at employer speech in U.S. labor law, compensation for banned handguns, and Cuba's expropriation of U.S. property. His intellectual property writings have examined biopiracy, a proposed trademark for Princess Diana, and copyright and access issues in the U.K. Higher Education Copying Accord. In 2001-2002 he wrote the research study for the U.K. Commission in Intellectual Property Rights on copyright issues in developing countries (see study paper no.5 at http://www.iprcommission.org/graphic/documents/study_papers.htm) and was a cochair of WIPOUT, the international IP counter-essay contest (see http://www.uea.ac.uk/~j013/wipout/index.html). He is now starting research for a book on the economics and politics of intellectual property. Professor Story has an LL.B. (Osgoode Hall Law School, Toronto) and an LL.M. (Cornell Law School).

Peter N. Weiss began work with the Strategic Planning and Policy Office of the U.S. National Weather Service of the National Oceanic and Atmospheric Administration in March 2000. His responsibilities include domestic and international data policy issues, with a view towards fostering a healthy public/private partnership. Mr. Weiss was a senior policy analyst/attorney in the Office of Information and Regulatory Affairs, Office of Management and Budget from 1991 to 2000. Mr. Weiss analyzed policy and legal issues involving information resources and information technology management, with particular emphasis on electronic data interchange and electronic commerce. He is primary author of the information policy sections of OMB Circular No. A-130, "Management of Federal Information Resources," and was a member of the Administration's Electronic Commerce Working Group. (See "A Framework for Global Electronic Commerce.") From 1990 to 1991 Mr. Weiss was deputy associate administrator for procurement law, Office of Federal Procurement Policy. In this position, he analyzed legal and policy issues affecting the procurement process. Major projects included examination of legal and regulatory issues involving procurement automation, policies, and FAR revisions to facilitate EDI, as well as ADP procurement legal and policy issues. From 1985 to 1990, Mr. Weiss was the assistant chief counsel for procurement and regulatory policy, Office of Advocacy, U.S. Small Business Administration. From 1981 to 1985, Mr. Weiss was in private practice in Washington, D.C. Mr. Weiss holds a B.A. from Columbia University and a J.D.

from the Catholic University of America, Columbus School of Law. A recent publication is "International Information Policy in Conflict: Open and Unrestricted Access versus Government Commercialization," in *Borders in Cyberspace*, Kahin and Nesson, eds., MIT Press, 1997.

Nico E. Willemse is a data management consultant for the InfoCom Project under the Environmental Information Systems Unit of the Directorate of Environmental Affairs, Ministry of Environment and Tourism of Namibia. He worked for two years as the principal training coordinator for the ACP-EU Fisheries and Biodiversity Management Project in southern Africa. He has a B.Sc. in zoology from the University of Namibia, an M.Sc. in international fisheries management from the Norwegian College of Fishery Science at the University of Tromsø, and is currently enrolled in an M.B.A. program at the University of South Africa. His research includes reconstruction of a time series of marine fisheries catches for Namibia, 1950-2000; major trends in the Namibian marine fisheries, 1950-2000; interpretation of marine fisheries catches for a 50-year time series; the "fishing down marine food webs" phenomenon in Namibia; and economic interpretation of a 50-year time series of marine catches off Namibia.

Mikhail Zgurovsky is rector of National Technical University of Ukraine's "Kiev Polytechnic Institute," and director of the Institute for Applied System Analysis of the National Academy of Sciences and Ministry of Education and Science of Ukraine. He is a Doctor of technical sciences and academician of National Academy of Sciences of Ukraine. Dr. Zgurovsky is a member of the Board on Science and Scientific-Technical Policy, head of the Association of Rectors of Technical Universities of Ukraine, and co-head of the Ukrainian Union of Scientists and Engineers. During 1994-1999, Dr. Zgurovsky was minister of Education of Ukraine and took an active part in working out a number of laws on education. He is a well-known scientist in the field of mathematics and cybernetics. His scientific interests and research embrace methodology of system analysis, the theory of decision making under uncertainty conditions, and analysis and modeling of various complex systems. He is a member of academies of sciences in many countries, senior member of IEEE, member of the Governing Board of UNESCO's Institute for Information Technologies in Education (Moscow, Russia), the national representative of Ukraine for CODATA, and a member of EDNES.

STEERING COMMITTEE MEMBERS

M. G. K Menon (*chair*) has had a distinguished career as a scientist and policy maker and has held a number of prominent appointments including secretary, Department of Science and Technology, and secretary, Department of Electronics, both in the Government of India; member, Planning Commission; and scientific adviser to the Prime Minister. He served as president of ICSU from 1988 to 1993. He is the recipient of such prestigious awards as Padma Bhushan and Padma Vibhushan in recognition of his distinguished service. He has also been honored with the Abdus Salaam award. He has been a member of Parliament in the Rajya Sabha. He has a Ph.D. from Bristol and was educated at Agra and Bombay. He is the president of Leadership for the Environment and Development in India

Carlos Correa is the director of the master program on science and technology policy and management at the University of Buenos Aires. He serves as a consultant in the fields of science and technology and intellectual property to many regional and international organizations, including UNCTAD, UNIDO, WHO, FAO, and the Inter-American Development Bank. He served as undersecretary of state for informatics and development in the Argentine national government from 1984 to 1989 and as the director of the U.N. Development Programme's Regional Programme on Informatics and Microelectronics for Latin America and the Caribbean from 1990 to 1995. He is a member of the Scientific Resource Group on Globalization of WHO and of the International Economics Law Association, and served as a member of the United Kingdom's International Commission on Intellectual Property Rights.

Dialo Diop is a lecturer in microbiology at the Université Cheikh Anta Diop in Dakar, Senegal. He received his M.D. from CHU Pitié-Salpêtrière in Paris. After working as a general practitioner in Senegal he returned to Paris to receive additional training in molecular virology. From 2000 to 2001 he served as the cabinet director for the Ministry of Higher Education and Scientific Research. He is currently in a Ph.D. program in medical molecular genetics at the Université Pierre et Marie Curie in Paris. Dr. Diop served as secretary-general of CODATA Senegal.

Farouk El-Baz is a research professor and the director of the Center for Remote Sensing at Boston University. He is also an adjunct professor of geology at the Faculty of Science, Ain Shams University, Cairo, Egypt. He received a B.Sc. in chemistry and geology from Ain Shams University, an M.S. degree in geology from the Missouri School of Mines and Metallurgy, and a Ph.D. in geology from the University of Missouri. He was elected a fellow of the Third World Academy of Sciences in 1985, and became a member of its Council in 1997. He represents that academy at the Non-Governmental Unit of the Economic and Social Council of the United Nations. Dr. El-Baz is a member of the U.S. National Academy of Engineering. He serves on the Board of Trustees of the new Alexandria Library, the Arab Science and Technology Foundation, the Egyptian Center for Economic Studies, the Egyptian-American Affairs Council, the Moroccan-American Council, the World Affairs Council of Boston, as well as the editorial boards of several international professional journals. He is a member of many national and international professional societies and a fellow of the Geological Society of America, the American Association for the Advancement of Science, the Royal Astronomical Society, and the Explorers Club. Dr. El-Baz has won numerous honors and awards, including NASA's Apollo Achievement Award, Exceptional Scientific Achievement Medal, and Special Recognition Award; the University of Missouri Alumni Achievement Award for Extraordinary Scientific Accomplishments; the Certificate of Merit of the World Aerospace Education Organization; the Golden Door Award of the International Institute of Boston; and the Award for Public Understanding of Science and Technology of the American Association for the Advancement of Science.

Dominique Foray is currently the directeur de recherche at the Centre National de la Recherche Scientifique (CNRS), a professor at the University of Paris-Dauphine, and a part-time member of the International Institute for Applied System Analysis (IIASA-Laxenburg bei Vienna). He received his Ph.D. in 1984 and his habilitation in 1992 from the Université Lumière of Lyon. In 1985 he joined the CNRS as a research fellow. In 1990 he joined the École Centrale Paris as professor of economics and returned to CNRS in 1994. From 1993 to 1995 he was a permanent consultant (part-time) at the OECD (Division for Science, Technology and Industry), where he contributed to the program on National Systems of Innovation. He received the distinction of outstanding research 1993 from CNRS. He was elected a research fellow at the ICER foundation (Italy) for the academic year 1999 and at the Institute for Advanced Study Berlin for the academic year 2000. His research interests include the economics of science and technology, the economics of production and distribution of knowledge, the exploration of the tension between diversity and standardization in the past and in the present, and the analysis of path-dependent processes of economic change.

Alexei Gvishiani is the director of the Centre of Geophysical Data Studies and Telematics Applications in the Russian Academy of Sciences and a professor of mathematics at Moscow State Lomonosov University. Dr. Gvishiani holds a Ph.D. in mathematics from Moscow State Lomonosov University and a doctor of sciences from Moscow Schmidt Institute of Physics of the Earth. Since 1994 he has been the vice-president of the European-Mediterranean Seismological Centre. His areas of scientific interest include artificial intelligence and applied mathematics in applications related to geophysical data acquisition, processing, and analysis, as well as the Internet and telematics applications for science, education, and the environment. Professor Gvishiani was recently elected vice-president of CODATA.

Elizabeth Longworth, principal of Longworth Associates, is a specialist adviser on dispute resolution, information issues, and digital technologies. She is also a specialist on the legal and ethical implications of information technology, the Internet, electronic commerce, electronic banking, international trade finance, and telecommuni-

cations. Following her admission to the bar as a barrister and solicitor and development of a practice in litigation, Ms. Longworth spent three years in Canada working in information law. Before setting up her own firm in 1991 she practiced in a large corporate law firm in both Wellington and Auckland, New Zealand, in her specialty fields. Ms. Longworth is the independent chair of the New Zealand telecommunication industry's self-regulatory body on number administration. She was the New Zealand nominee to the UNESCO meetings in Seoul and Monte Carlo in 1998 and is a member of the Sub-Commission on Communication of the New Zealand National Commission of UNESCO. She continues to work for UNESCO (Paris) on the digital divide. Ms. Longworth is the author of a report on cyberspace law, published in New Zealand in 1998 and currently in press by UNESCO (Paris), and of the leading text on New Zealand's privacy laws. Ms. Longworth graduated with a bachelor of laws from Victoria University of Wellington, New Zealand, and a master of laws from Osgoode Hall, York University, Canada.

Lulama Makhubela is the manager of research information for the South African National Research Foundation. Earlier she was a senior lecturer in the Department of Library and Information Science at the University of the Western Cape. She earned her master's degree in librarianship at the College of Librarianship, Wales, and her Ph.D. in library and information science from the University of the Western Cape. She is a member of the South African CODATA Committee.

Erik Sandewall is a professor of computer science and vice-rector at Linköping University, Sweden. Dr. Sandewall is a member of the Royal Swedish Academy of Sciences, the Royal Swedish Academy of Engineering Sciences, and a fellow of the American Association for Artificial Intelligence. He received his Ph.D. from Uppsala University, after spending a year of doctoral studies at the Stanford Artificial Intelligence Laboratory. He has also been a visiting associate professor at the Massachusetts Institute of Technology Artificial Intelligence Laboratory. He was recently appointed as the chair of the ICSU Press.

Mary Waltham is an independent consultant. She was most recently the president and publisher of *Nature* and the Nature family of journals in the United States, and formerly the managing director and publisher of *The Lancet* in the United Kingdom. She founded her own consulting company two years ago. Its purpose is to help international scientific, technical, and medical publishers to confront the rapid change that the networked economy poses to their traditional business models, and to develop the new opportunities to build publications that deliver outstanding scientific and economic value. Ms. Waltham has worked at a senior level in science and medical publishing companies across a range of media, which include textbooks, magazines, newsletters, journals, and open learning materials. She served on the National Research Council's Committee on Community Standards for Sharing Publication-Related Data and Materials and the Steering Committee for the Symposium on Electronic Scientific, Technical, and Medical Journal Publishing and Its Implications.

Ferris Webster is a professor of oceanography in the College of Marine Studies of the University of Delaware. He received his B.Sc. and M.Sc. degrees in physics at the University of Alberta and his Ph.D. in geophysics at the Massachusetts Institute of Technology. Beginning at the Woods Hole Oceanographic Institution he has held a number of scientific positions, becoming senior scientist, chairman of the Physical Oceanography Department, and associate director for research. During this period he spent a sabbatical year at the National Institute of Oceanography in England. Between 1978 and 1982 Dr. Webster served as assistant administrator for research and development of the National Oceanic and Atmospheric Administration. In 1982 he became a senior fellow with the National Research Council. He joined the University of Delaware in 1983, where he serves as the director of the Oceanography Program. His research interests include the role of the ocean in climate change, ocean variability, time series analysis, and oceanographic data management and processing. Since 1994 he has served as the chairman of the Panel on World Data Centers of the International Council for Science. He is also the chair of the ICSU-CODATA ad hoc Working Group on Access to Data and Information, and served as chair of the National Research Council's Committee on Geophysical and Environmental Data.

Appendix C

Symposium Attendees

Minella Alarcon
UNESCO
France

Ali Al-Mashat
UNESCO
France

William Anderson
Praxis101
United States

Francis Andre
INIST-CNRS
France

Jim Ashling
Ashling Consulting
United Kingdom

Paulo Henrique Assis Santana
CNPQ
Brazil

Chrisanthi Avgerou
London School of Economics
United Kingdom

Andrew Badran
Philadelphia University
Jordan

Enric Banda
European Science Foundation
France

Avril Bernard
University of Bergen
Norway

R. Stephen Berry
University of Chicago
United States

Massey Beveridge
University of Toronto
Canada

Kwasi Boakye-Akyeampong
Computer Professionals for Social Responsibility
United Kingdom

Francoise Bousquet
AKELA
France

Aileen Boyd-Squires
Blackwell Publishing
France

Gilberto Câmara
Instituto Nacional de Pesquisas Espaciais
Brazil

Anne Carblanc
OECD
France

Kathleen Cass
CODATA
France

Hilda Cerdeira
The Abdus Salam International Centre for Theoretical Physics
Italy

Leslie Chan
Bioline International
Canada

Swayandipta Pal Chaudhuri
Center for Communications and Information Research
India

Philippe Chevet
Maison des Sciences de l'homme
France

Liu Chuang
Institute of Geography and Natural Resources
China

Robin Cowan
University of Maastricht
The Netherlands

David Dickson
SciDev.Net
United Kingdom

Dialo Diop
Unite de Biochimie Cellulaire
France

Thomas Dreier
University of Karlsruhe
Germany

Sarah Durrant
International Network for the Availability of Scientific Publications
United Kingdom

Jacob Eduard
Society of Internet in Medicine
Holland

Farouk El-Baz
Boston University
United States

Julie Esanu
The National Academies
United States

Dominique Foray
Centre National de la Recherche Scientifique (CNRS)
France

Clemente Forero-Pineda
Universidades de los Andes y el Rosario
Columbia

Frederick Friend
Joint Information Systems
United Kingdom

Jean Garnier
INRA/NIH
France

Ruggero Gilyaresky
Moscow State University
Russia

Dov Greenbaum
Yale University
United States

Jane Bortnick Griffith
National Library of Medicine
United States

Robin Gross
IP Justice
United States

Francoise Guillaume
NIRA Rennes
France

Alexei Gvishiani
Russian Academy of Sciences
Russia

Shuichi Iwata
University of Tokyo
Japan

Andrew Kaniki
National Research Foundation
South Africa

Stephen Katz
Food and Agriculture Organisation of the United Nations
France

Izida Khamidoullina
LENTIC
Belgium

Milan Konecny
Masaryk University
Czech Republic

Subhash Kuvelker
Kuvelker Law Firm
United States

Nathalie Larmanjat
Blackwell Publishing
France

Elizabeth Longworth
Industry New Zealand
New Zealand

Michel Loots
World Information Transfer
Belgium

Murali Krishna Lyyanki
JNT University
India

Gudrun Maass
OECD
France

Barry Mahon
International Council for Scientific and Technical Information
France

Lulama Makhubela
National Research Foundation
South Africa

Diego Malpede
Third World Academy of Sciences
Italy

Elisabetta Marinoni
Universita di Padova
Italy

Tony Marjoram
Division of Basic and Engineering Sciences
France

Koïchiro Matsuura
UNESCO
France

Claudio Menezes
UNESCO
France

M. G. K. Menon
LEAD India
India

Victor Montviloff
UNESCO
France

Pascal Morand
WSIS Swiss Secretariat
Switzerland

Perumalsamy Navaraj
Yadava College
India

Ndaendelao (Emma) Noongo
Directorate of Environmental Affairs
Namibia

Jean-max Noyer
Univerite Paris 7
France

Raphael Ntambue
CNRS
France

Harlan Onsrud
University of Maine
United States

Nicole Pinhas
INSERM
France

Jean Luc Poncelet
PAHO Emergency Preparedness Program
United States

Mukund Rao
Earth Observation Systems
India

Agnes Raymond-Denise
Institut Pasteur
France

Gabriel Regus
UNESCO
France

Jerome Reichman
Duke University Law School
United States

Jonathan Robin
Internet Society European Chapters Coordination Council
France

John Rose
UNESCO
France

Steve Rossouw
Cape Technikon & Technikon SA
South Africa

Jean Jacques Royer
Centre de Recherches Pétrographiques et Géochimiques
France

D. K. Sahu
JPM
India

Kumar Sahu
Ecole Nationale des Ponts et Chaussees
France

Marjut Salokannel
University of Helsinki
Finland

Erik Sandewall
Linköping University
Sweden

Susan Schneegans
UNESCO
France

Peter Schröder
Ministry of Education
Holland

Charles Schweick
University of Massachusetts
United States

Laurents Sesink
Netherlands Institute for Scientific Information Services
Holland

Dennis Shaw
University of Oxford
United Kingdom

Anzhelika Sineok
Rostov State University
Russia

Sylvie Sou
Biblioteque J Hamburger
France

Charles Spillane
Institute of Plant Biology
Switzerland

Alan Story
University of Kent
United Kingdom

Atago Takaharu
Japan Science and Technology Corporation, Paris Office
France

Christophe Tuffery
ESRI France
France

Paul Uhlir
The National Academies
United States

Paul Van Brandt
Univ. Cath. de Louvain
Belgium

Geert Van Grootel
Ministry of Flanders, Science Division
Belgium

Janneke Van Kerson
Dutch Digital Heritage Association
Holland

Giuseppe Vitiello
Institute for Security Studies
France

Mary Waltham
Publishing Consultant
United States

John Webb
UNESCO
France

Ferris Webster
University of Delaware
United States

Peter Weiss
U.S. National Weather Service
United States

Nico Willemse
Ministry of Environment and Tourism
Namibia

Wendy White
The National Academies
United States

Gordon Wood
National Research Council
Canada

Paul Wouters
NERDI
Holland

Dong Wu
UNCTAD
Switzerland

Zelalem Wudeneh
Ethiopian S&T Commission
Ethiopia

Enrique Wulf-Barreiro
CSIC
Spain

Baoping Yan
Computer Information Center
China

Michail Zgurovsky
University of Ukraine
Ukraine

Marie-Claude Zikra
National Center of Economic Research
France

Dieter Zinnbauer
London School of Economics
United Kingdom

Appendix D

Acronyms and Initialisms

AJOL	African Journals OnLine
arXiv	e-print service
ASEA	Association of Surgeons of East Africa
BMJ	*British Medical Journal*
CERN	European Organisation for Nuclear Research
CIS	Commonwealth of Independent States
CODATA	Committee on Data for Science and Technology
CRIA	Reference Center for Environmental Information (Brazil)
CRID	Regional Disaster Information Center
DATAD	Database of African Theses and Dissertations
DMCA	Digital Millennium Copyright Act, United States
DRM	digital rights management
EDNES	Earth Data Network for Education and Scientific Exchange
eSAP	electronic Supply of Academic Publications
E.U.	European Union
FOIA	Freedom of Information Act
GEANT	detector description and simulation tool
GIS	geographical information systems
GNU	GNU not Unix (refers to Stallman's free software movement)
GPS	Global Positioning System
HBU	historically black university
HINARI	Health InterNetwork Access to Research Initiative
HWU	historically white university

HWU-A	historically white Afrikaans university
HWU-E	historically white English university
ICSTI	International Council for Scientific and Technical Information
ICSU	International Council for Science
ICT	information and communication technology
INASP	International Network for the Availability of Scientific Publications
INDMED	Indian Biomedical Journals Database
INSERM	French Institute of Health and Medical Research
IP	intellectual property
IPR	intellectual property right
ISTIP	International Scientific and Technical Information Programs
JPM	*Journal of Postgraduate Medicine*
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	normalized difference vegetation index
NEPAD	New Partnership for Africa's Development
NGO	nongovernmental organization
NIST	National Institute for Science and Technology, United States
NLM	National Library of Medicine, United States
NSDI	National Spatial Data Infrastructure
OAI	Open Archives Initiative
OECD	Organisation for Economic Co-operation and Development
OSVITA	Ukrainian national educational information system
PAHO	Pan American Health Organization
PDB	Protein Data Bank
PERI	Programme for the Enhancement of Research Information
R&D	research and development
SADC	South African Development Community
SciDev.Net	Science and Development Network
SPARC	Scholarly Publishing and Academic Resources Coalition
STACCIS	Support for Telematics Applications Cooperation with the CIS
TPM	technological protection measures
TRIPS	Trade-Related Aspects of Intellectual Property (agreement)
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNIDO	United Nations Industrial Development Organization
URAN	Ukrainian Research and Academic Network
WHO	World Health Organization
WIPO	World Intellectual Property Organization
WSIS	World Summit on the Information Society
WTO	World Trade Organization
XML	extensible markup language